

Infectious diseases are not banished

The US Institute of Medicine has just produced a document that is either scary or a spur that will persuade governments to pay more attention to old-fashioned public health.

LULLED by unfounded belief in the power of antibiotics to make infectious diseases things of the past, Western governments have been making political decisions to reduce research and disease surveillance. They may yet be very sorry. The unexpected emergence of AIDS should have taught them an important lesson — that viruses, harboured by animals, may well infect people who push their residential way into previously uninhabited forests or fields. But it is easy to dismiss the human immunodeficiency virus (HIV) as a special case. It tends to infect readily identifiable (and often socially estranged) groups, such as drug addicts who share dirty needles, and it is transmitted only with relative difficulty. You do not catch AIDS when someone sneezes.

But what about tuberculosis (TB)? TB is now on the rise in the United States, including a form with a multiple drug resistance, having gained a stronghold among people who are immunosuppressed (AIDS patients, for instance) and those living in poverty without good medical care. TB is something you can catch when someone sneezes, or coughs. Are public health officials (and the politicians who control their budgets) ready for the very real possibility that TB could again become epidemic in the United States and Europe?

The answer, at present, is no. That at least is the opinion of a group convened by the US Institute of Medicine to study the threat of emerging microorganisms. The group's leaders, Joshua Lederberg of the Rockefeller University and Robert Shope, director of the Arbovirus Research Unit at Yale University, hope that the menacing data in their report, *Emerging infections*, will have a galvanizing effect. Shope believes that the world is still vulnerable to something along the lines of the 1918-1919 influenza pandemic, which killed 20 million people worldwide. "It's happened once", he says: "it could happen again."

The evidence, alas, is compelling. Diseases once thought to be on the wane or under control gather new vigour. Malaria (both *Plasmodium falciparum* and *Plasmodium vivax*) is increasingly resistant to chloroquine and there are data suggesting newly acquired resistance to mefloquine, on the market only a few years. Lyme disease, caused by *Borrelia burgdorferi* (a spirochete) and transmitted by ticks (Ixodes) that live on deer, was virtually unknown a decade ago, but is now the most common vector-borne disease in the United States and an increasing threat in Europe.

The litany of microbial threats to human populations that

were once controlled or localized seems endless. Rift Valley fever, Hantaan virus, and yellow fever are among them. Worse, mosquitoes and other vectors, once eradicable by pesticides, are often now resistant. And a previously neglected bacterium *Helicobacter pylori* is now believed to be a cause of peptic ulcers, raising the worry of infectious cause of diseases previously believed to be functional. Where will it end?

The Lederberg-Shope committee takes the view that the genetic variability of viruses such as those responsible for influenza and AIDS is not the only cause for alarm. The resistance of microorganisms to known antimicrobial drugs is another. But so, interestingly, are changes in human behaviour and living patterns. Lyme disease is a vivid example. When the forests on the East Coast of the United States were cleared for farming in the nineteenth century, the habitat for the vector tick was destroyed. But reforestation early this century and the movement of populations from cities to wooded suburbs brought people, deer and ticks, and hence the Lyme spirochete, too close together. Foreign travel is now a further exacerbation of the problem.

What kinds of risks does the committee have in mind? It is specific on this issue, raising the question of a hypothetical outbreak of yellow fever in New Orleans, a city of 500,000 people. *Aedes aegypti*, which carries yellow fever, is known in New Orleans, as is *Aedes albopictus*, a newer vector. Suppose yellow fever were to break out. The available stocks of vaccine would be depleted within a few days and new supplies would have to come from Brazil, which alone has large stocks. But what the committee calls "acceptable" (or environmentally correct) pesticides to kill the mosquito vectors are not available. The committee predicts that 100,000 people would go down with yellow fever and that 10,000 would probably die within 90 days. That is a gloomy prospect for a rich city in a rich country.

Must it take such a catastrophe to renew governments' commitments to avert the threat of emerging microorganisms? Throughout the world, but particularly in the United States and Europe, money for research and for disease surveillance and reporting systems has been cut during the past two complacent decades. That is why TB is reemergent. Yet few problems are as difficult to deal with as those whose solutions are known but whose threat seems distant. Or, as Lederberg has observed, when there is a direct correlation between funding and special interest lobbying,



who will defend the long-term and general public interest? If the Institute of Medicine has reminded politicians of what many suppose to be their special role, it will have done a power of good. □

Market ideology

The British government seems to believe that a market in energy absolves it from responsibility.

THE hapless British government, still punch-drunk from the expulsion of sterling from the European Exchange Rate Mechanism and committed to fighting the Maastricht Treaty through the British Parliament (from 10 November) without knowing where the votes will come from, last week engineered a third crisis for itself by sacking 30,000 coal miners. More accurately, the government accepted the recommendation of British Coal, the still-nationalized rump of the coal industry, that 31 collieries should be closed by the end of the financial year in March, some of them immediately, and also agreed to find an extra £1.2 billion to make redundancy payments to the people who would lose their jobs. And then the government casually made the plan public, three days before some of the collieries concerned were shut, without apparently expecting the ructions its casualness would evoke. It remains to be seen whether the government will have been able to modify the plan sufficiently to save its political skin on Wednesday, when the issue will have been debated in the House of Commons.

Whether or not the government survives, the underlying technical issues will not go away. They have to do with the proper role of government in decisions about the use of different sources of energy. In Britain, the issue is complicated by the privatization two years ago of the electricity industry, British Coal's chief customer, which now consists of four companies (one Scottish) which generate electricity and a dozen distribution companies (in England and Wales) which sell electricity and also have the right to generate their own. The crisis in the coal industry has come about because the distribution companies, resentful of their potentially costly dependence on the generators, have set about building generating capacity based on natural gas. The generators, wishing not to be left behind, have done the same. Their combined demand for coal has as a consequence collapsed.

The government's line is that this outcome has been dictated by the market for bulk fuel, and that its own responsibility is limited to dealing with the consequences for British Coal. This line would be more persuasive if the market in bulk fuel were free and efficient in the economists' sense. But it is not. One of the generating companies, Nuclear Electric, exists because the government rightly took the view, at the time of privatization, that it would be absurd to scrap Britain's nuclear power stations, which generate some 18 per cent of total electricity, simply because private investors would not readily subscribe for shares in companies likely to be faced with indeterminate

but large decommissioning costs. Nuclear Electric thus remains the property of the government, but is financed by electricity consumers through a levy on the cost of sales. In short, Britain's energy market is deliberately biased for a strategic reason.

Moreover, the market is far from free. Most of those who use it are monopolies of one kind or another. That is substantially true of the electricity distribution companies (but the generators have the right to sell directly to large consumers, which takes some icing off the cake). So are many of the fuel suppliers, including British Coal (which has a *lein* on British coal deposits) and British Gas, the privatised version of the previously nationalized industry, most of whose business now consists of buying natural gas from oil companies with interests in the North Sea and selling it on, at a profit, to domestic and commercial users. The "dash for gas" among electricity generators in the past few months has come about because oil companies with wells in the North Sea have won the right to sell direct to electricity generators. Nobody knows to what extent the attractiveness of gas stems from suppliers' anxiety to increase market share, or simply to get their money back more quickly.

One irony of last week's decision is its timing. A bill allowing the government to sell off the coal industry is already in the British Parliament, but will now be held up until the row about the 30,000 miners has been settled. So why, instead of telling 30,000 miners that there would now be no work for them, did not the government at least offer them the option of owning the mines from which they are to be expelled? That would at least have been a market solution to what has been described as a market problem. Could it be that the British government fears that the remainder of the industry would be less saleable if there were pre-existing competition?

The plain truth is that even when the government has rightly shed direct responsibility for the management of the energy industries, it will still have a continuing responsibility for the ways in which energy is used. The existence of Nuclear Electricity shows that it accepts the responsibility in respect of nuclear power. The heavy excise taxes on motor fuel similarly imply a mixture of sentiments and calculations about energy conservation, traffic congestion and the cost of building highways. It would be in no way inconsistent with the way the market has been operated so far that there should be a tax on power-station gas at least until the costs of closing half the coal industry have been covered.

None of that, of course, bears on the reasons why the British government roused such anger with its decision last week. Rather, the precipitate firing of 30,000 people, with only newspaper leaks for warning, was rightly regarded as callous, while having done so in the middle of a long recession was plainly foolish. What the British government should learn from this experience is that the existence of a market does not absolve governments from responsibility. That is especially so when they make the rules — and break them when they choose. □

Drastic changes called for in French space programme

Paris. French space policy is badly managed and lacks direction, and its four-fold growth in the past decade threatens the health of other areas of research. That is the conclusion of a report issued earlier this month by the Comité National d'Évaluation de la Recherche (CNER), formed to evaluate the country's research efforts. It says that what is needed is a parliamentary debate on space policy, greater emphasis on research, better financial and scientific management of existing programmes and increased accountability to the scientific community.

The crisis in French space, the world's third largest programme and the driving force behind European efforts in the field, stems from a decision by the French space agency Centre National d'Études Spatiales (CNES) to include manned space flights and a space station in its programme. That decision helped its budget to grow from FF2.1 billion (US\$400 million) in 1982 to FF9.2 billion next year, but it left little for other agency programmes. Similarly, CNES's increasing share of the civilian research budget — from 6.4 per cent in 1982 to 18 per cent next year — has squeezed out other worthy projects.

The problem may be alleviated next month if the European Space Agency, as is expected at a ministerial meeting in Granada, Spain, abandons its plans for an independent European space programme. But CNER is equally concerned about the origins of the problem, and how to prevent it from recurring.

The committee, created in 1989 by France's research minister, Hubert Curien, emphasizes the government's lack of control over space policy and the agency's poor handling of its finances and its research portfolio. Despite being the largest item in the research budget, space research has lacked a coherent strategy. Important decisions are made at various levels within the government, and two supervisory bodies formed in 1988 to improve the situation lack clear terms of reference and rarely meet.

Among other deficiencies, the committee could find no written record of the country's decision in 1985 to build Hermès, for instance, and a programme conceived as a single research satellite, SPOT 1, has somehow grown into a series of four commercial satellites. The SPOT programme is still being funded through the research budget and, despite costing FF8 billion, "nobody knows if the government ever wanted it".

One necessary change would be to form

a powerful space committee within the government to bring CNES under executive control. A second remedy would be simpler accounting procedures allowing the agency's costs to be compared with the rest of civilian research spending. Specifically, the committee wants to end the practice of using the civilian research budget to fund such items as the METEOSAT satellite and road

systems, computing, artificial intelligence, robotics, electronics and the automation of space activities. The committee also questions the wisdom of going ahead with a heavy solid-fuel booster rocket, Ariane V (FF2.7 billion in 1993), designed chiefly to serve the Hermès project, and recommends instead that CNES revives the Ariane IV rocket, which it believes is better suited for both scientific and commercial launches.

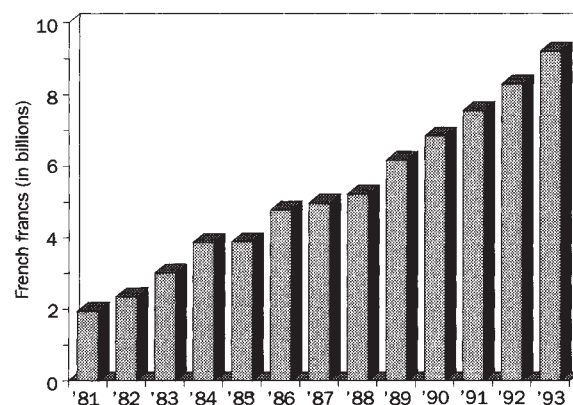
Although Ariane is a notable success, CNES has also had many failures. Facing competition from France Telecom in 1988, CNES pressured companies and banks into supporting a risky venture to develop a commercial position locator system called LOCSTAR. The company went out of business last year after losing FF700 million.

The report also suggests that CNES reduces its industrial activities and return to its research roots. Towards that end, the committee recommends that CNES relinquishes its role in mature telecommunications and meteorological projects to commercial companies and give up its one-third share in Arianespace.

The evaluation committee has no power to enforce its recommendations but it intends to pursue the issues raised in its report. Neither CNES nor the Ministry of Research and Space will comment, but Christian Bécle, president of CNER, says that the report's findings have "moved" the research ministry and that changes are inevitable.

Declan Butler

French space spending takes off



Source: French Ministry of Research and Space

construction at the agency's launch centre in Guyana.

The committee criticizes the agency for failing to recruit new talent to offset an ageing research staff and being too willing to follow the lead of the United States with such projects as Hermès and Columbus. It recommends that CNES pays greater attention to such technologies as new launch

NASA carves up science programme

Washington. The administrator of the US National Aeronautics and Space Administration (NASA), Daniel Goldin, took the science community by surprise last week by splitting the space agency's science office in two and shunting aside its popular chief, Lennard Fisk.

The upheaval is part of Goldin's effort to lower costs, improve quality and reduce the time it takes to launch spacecraft (see *Nature* 358, 701; 1992). He also wants to bring civilian-minded NASA closer to the military, which is eager to place its sophisticated sensors aboard the agency's science satellites.

The reshuffling is more than rearranging the deck chairs on the ship of a sinking Republican administration. While big

changes have been expected since Goldin took over in April, the reorganization raises more questions than it answers — and the confusion is not likely to subside until next year. If George Bush loses the election, Goldin's plans could evaporate if Bill Clinton wants a new NASA administrator. But Clinton could also decide first to see if Goldin's plans succeed.

One uncertainty is that of where NASA's life and materials sciences programmes fit in this regime. Understanding how people function in space is now NASA's primary justification for the space station, and manufacturing flawless crystals and other materials in space is said to be the major industrial application of the station. Yet these two areas are yet to be reassigned. They could

find a home either within the new office devoted to Mission to Planet Earth or fall under the programme for planetary science and astrophysics, the two successors to the space science and applications office that Fisk has led since 1987.

Fisk, having lost control of a \$2-billion programme, has been assigned to the administrator's office as the agency's first chief scientist. At a press conference on 15 October, Fisk said that his new position demonstrates science is vital to NASA, but it is more likely that Fisk has been kicked upstairs.

A familiar sight on Capitol Hill, Fisk gets on well with the congressional committees that oversee NASA. A former professor of physics at the University of New Hampshire, Fisk has won grudging support from researchers for his long-term plan for the agency's space science efforts. His office is responsible for such projects as the multibillion-dollar Earth Observing System (a constellation of satellites to monitor the environment), the Advanced X-Ray Astrophysics Facility to be launched later in the decade and the near-sighted Hubble Space Telescope.

In recent years, however, the White House National Space Council has been pressing NASA to buy the "cheaper, faster, better" approach favoured by the vice president, Dan Quayle. Fisk has been a reluctant supporter of the concept.

Goldin took office after Richard Truly was fired on 10 February for failing to follow Quayle's orders and immediately began pressing his managers to shrink their programmes. Goldin is eager to bring in his own team, and removing Fisk sees a natural step in that process.

The pieces of the old science office fall — at least temporarily — to Shelby Tilford, director of Earth sciences, and Wes Huntress, director of Solar Systems exploration. But Goldin promised last week to allow other NASA employees and outsiders to compete for those two jobs and a host of others. The move could also lower barriers to cooperation with the Defense and Energy departments, whose scientists now want to use sophisticated sensors designed for the Cold War for civilian missions.

Another victim of the reorganization is the office of aeronautics and space technology. Its head, Pete Peterson, has been asked to conduct a long-term study of US aeronautics and space facilities, while Cecil Rosen, who was aeronautics director, takes over as acting chief of a strengthened aeronautics office.

Goldin has transformed the space technology portion of the agency into an advanced concepts office run by Gregory Reck that will also include the agency's commercial programmes. Courtney Stadd, formerly of the White House National Space Council, assumes the job of acting deputy associate administrator for the new office.

Andrew Lawler

Racial tensions entangle NIH in dispute over AIDS drug

Washington. Under pressure from the Nation of Islam and other African-American activist and political groups, the National Institutes of Health (NIH) is reconsidering a report issued earlier this year rejecting Kemron and other interferon alpha-based drugs as useful treatments for people with AIDS. NIH officials concede that they were politically naive when they evaluated the drugs — often touted as an 'African AIDS cure' because some early clinical trials were conducted in Kenya — in the same way as any other experimental therapy without accounting for the desperation and suspicion of those affected by the AIDS epidemic.

Next week, advocates of oral interferon alpha will hold a long-awaited meeting with NIH officials as the first step towards NIH-sponsored clinical trials of the drug for those with AIDS. But even this is fraught with racial tensions. Because many African-American groups believe that the government is ignoring — or even encouraging — the problem of AIDS within minority populations, NIH asked the National Medical Association (NMA), an organization comprised mostly of African-American physicians, to invite the participants and serve as the host. Yet in the same week that invitations were sent out, NIH mailed to NMA members a copy of its April report rejecting interferon alpha as a treatment for AIDS. Advocates of the drug accuse NIH of trying to prejudice the debate; NIH says that it was simply trying to disseminate relevant information to all those interested.

The agendas of the participants reflect the distance separating them: advocates of interferon alpha, who have been selling and distributing the drug to AIDS patients for more than a year, want NIH to evaluate their clinical data and sponsor more formal clinical trials, while NIH officials, including

Daniel Hoth, director of the NIH division of AIDS, have taken the position that they first need to learn more about the claims being made for the drug. However, NIH has agreed to examine some new data and conduct another assessment of interferon alpha to supersede its earlier report. And last month the Congressional Black Caucus, a group of African-American legislators, met NIH officials and argued the case for Kemron.

Various claims have been made for the efficacy of low-dose oral interferon alpha in treating AIDS. After the initial 1989 clinical trial in Kenya, Davy Koech, director of the Kenyan Medical Research Institute, reported that CD4 counts had improved substantially in many of the participants and that as many as 20 per cent had become HIV-seronegative. These results created a worldwide demand. However, since then even Koech has played down 'seroconversion' (which has not been replicated) as a measure of efficacy and has instead focused on such indicators as the level of symptoms and overall health, energy and appetite of the patients.

Although NIH continues to stand behind the April report (see *Nature* **356**, 648; 1992), its conclusions have been rejected by many of the groups who use and prescribe the drugs. Advocates point out that NIH focused on the dramatic seroconversion and CD4 claims, rather than the more subjective (and better replicated) reports of symptom reduction. Barbara Justice, a physician affiliated with the Nation of Islam who runs an AIDS clinic in New York, says that NIH also erred in favouring randomized and placebo-controlled clinical trials and immune system measures such as CD4 counts and should have focused instead on whether AIDS patients were healthier — by any measure — after interferon-alpha treatment. "NIH is hung up on 'double-

NHS opens meta-analysis centre

London. Britain's plan to coordinate and strengthen research by the National Health Service (NHS) takes an important step forward next month with the opening of a new centre for meta-analysis.

The Cochrane Centre in Oxford will compile and systematically review clinical and other randomised controlled trials on topics from drugs to surgical instruments. It will also help researchers to perform meta-analyses or overviews of specialist areas with the goal of making the data available to clinicians. The creation of the centre, with an annual budget of £300,000, indicates the increasing importance of meta-analysis in

interpreting clinical trials and is a response to criticism that the results are implemented slowly and inconsistently. The NHS is planning another facility for reviewing work other than randomized controlled trials and has already created a database that, for the first time, gives a clear picture of the amount and nature of research going on in the NHS.

Next month, the NHS is also expected to complete the hiring of 14 regional research and development directors responsible for implementing the new research strategy. The NHS has promised to spend 1.5 per cent of its budget on research by 1997, around £375 million at today's prices. **Ian Mundell**



African-American doctors like Barbara Justice (left) have stepped up pressure on Daniel Hoth (right) and other NIH officials for government-sponsored clinical trials of interferon alpha.

blind this' and 'T-Cell that', she says. "I'm hung up on my patients getting better."

Her argument reflects the problem of applying the usual methods of scientific assessment to an issue in which there is great hostility towards the government. NIH sponsored the analysis because it felt it had a public health obligation to determine if an unapproved drug being used by thousands of people actually worked. Yet when NIH concluded that it did not, it is not surprising that the agency failed to get that message across. The African-American groups who see interferon alpha as a African solution to the AIDS problem were not persuaded to drop their efforts by a government report that, to them, was based on obscure scientific standards rather than real experience. And some errors in the NIH report, including aggregating incompatible trials, fuelled suspicions that the US government was trying to suppress the drug.

Cell Culture Co. Inc, as a treatment for respiratory disease in cattle. Cummins has conducted clinical trials of interferon, with varying success, on a number of human and animal diseases. Earlier this year, he received approval for interferon-alpha clinical trials on AIDS patients in the United States.

In 1989, Cummins travelled to Kenya to test interferon alpha in cattle diseases. He met Koech, who was studying AIDS therapies and had seen a 1986 paper by Cummins on the use of low-dose interferon alpha to treat feline leukaemia, a disease that is related to AIDS. Cummins showed Koech data from a trial he had started the previous year in Texas of interferon on HIV-positive patients. He then provided Koech with interferon-alpha powder from Japan and Koech began uncontrolled clinical trials on Kenyan AIDS patients. After a few weeks Koech reported dramatic improvements, and announced at a press conference that he had

The twisted history of Kemron and a host of similar interferon-alpha-based products contributes to the confusion. Interferon-alpha therapy was developed 20 years ago by Joseph Cummins, a Texas veterinary microbiologist and president of the Amarillo

discovered a secret AIDS treatment he called KEO-89. In February 1990, he announced that the drug was oral interferon alpha and would be marketed as Kemron.

Hoping to replicate the results in a placebo-controlled trial, Cummins sent Koech interferon tablets in January 1990. But Koech decided that results were so promising that it would be unethical to withhold the drug from a placebo population.

The situation soon became chaotic. Cummins discovered that many of the interferon-alpha tablets — and the matching placebos — that he had shipped to Koech were instead being sold on the black market for as much as \$50 each.

Since then, Cummins' company has sued an Australian company for patent infringement after it began marketing a product called Immunex in Africa and the United States, and a US company, also called Immunex, has threatened to file a trademark infringement suit against the US importers. The result is that the drug will now sell in the United States as Immuviron, without FDA approval, at about \$8 a dose, nearly one hundred times the cost of production.

Advocates continue to use and prescribe oral interferon alpha, pointing out that, unlike accepted AIDS drugs such as AZT, it is nontoxic in low doses. Although it may not cure AIDS, they believe that it alleviates symptoms and want NIH to evaluate as it does other AIDS drugs — not as a cure, but as a therapy. Given the continuing demands for clinical trials from the hundreds of US AIDS patients being treated with interferon alpha who believe that their government is suppressing a promising treatment, NIH may find these arguments hard to resist.

Christopher Anderson

NSF drops flat-rate plan for grants in mathematics

Washington. A plan hastily conceived by the US National Science Foundation to award mathematics grants of a predetermined amount (see *Nature* 359, 94; 1992) has been withdrawn after complaints that it would not achieve the goal of funding more research. NSF had regarded the plan as a model for preserving the vitality of various disciplines without spending more by standardizing the size of grants and limiting the amount of overhead (so-called indirect costs) paid to universities for sponsoring research on campus.

Last May, NSF's mathematics advisory group asked the foundation to find a way to fund 100 or so more proposals each year despite a flat budget. The obvious solution was to reduce the size of existing grants by a small amount and to apply the money saved to new grants. By the end of August, NSF had a plan by which investigators would receive grants of either \$20,000 or \$30,000, depending on the quality of the research, and \$10,000 more for a graduate

or postdoctoral student. Budgets would not be negotiated individually, and no allowances would be made for variations in indirect costs and other expenses.

NSF sent a letter to the mathematics community explaining the plan and announcing that it would take effect on 1 October. But officials soon discovered that researchers were not pleased with its work. "They screwed up, there's no other way to put it", says Jerry Bona, former chairman of the advisory committee and a mathematics professor at Pennsylvania State University.

One problem with the NSF plan was that it would not have funded any more grants. "We would have wound up raising some grants and cutting others drastically", says Bona, "but there would have been no new grants in the first two years." Another problem was the approach to indirect costs. By failing to recognize the wide variation in rates among research universities, set during annual negotiations with the government, NSF was ignoring existing federal

policy and penalizing private universities, where costs are generally higher than at public universities.

After meeting the science policy committee of the American Mathematical Society in early September and hearing protests from the community, NSF decided to defer the plan. "NSF's policy is to pay the appropriate indirect costs and this project does not change that policy", says Judith Sunley, executive officer for the NSF directorate that includes mathematics and a former director of the division.

The mathematics advisory committee will discuss the issue at its meeting next week and is likely to ask NSF to devise a better plan. Sunley says that NSF hopes to have a new version ready by January, when Fred Wan of the University of Washington joins NSF as director of the mathematics division. Until then, NSF will continue its current policy of negotiating individual budgets for each grant.

Jeffrey Mervis

Clinton emphasizes technology over science in choosing advisers for presidential campaign

Washington. If a man is known by the company he keeps, Bill Clinton, the Democratic candidate for president, could be the most science-orientated leader in US history. Clinton has not only picked as his running mate Senator Al Gore (Democrat, Tennessee), arguably the most technologically informed member of that body and the most enthusiastic for science, but he has dozens of prominent experts on hand helping him to make science and technology an important plank in his platform.

But, attentive as Clinton appears to be towards science and technology, it would be a mistake to assume that this will translate into more money for basic research. Like the candidate, the people around him are focused more on technology than on science. Not a single practising — or even former — scientist is part of his inner circle.

Two influential figures are John Sculley and John Young, the presidents of Apple Computers and Hewlett-Packard, along with independent economic and business consultants Ira Magaziner and Tom Schneider. As a group, Clinton's advisers have focused on ways to strengthen the US economy, with science and technology a means to that end. When the candidate talks about shifting defence spending to the civilian sector and boosting the prominence of science and technology, what he really has in mind is developing products that will sell well.

Clinton's public statements reflect this concern about industry. He has advocated a civilian equivalent to the Defense Advanced Research Projects Agency to target and fund research aimed at important technologies and permanent tax credits for industrial research. National laboratories would set aside as much as 20 per cent of their budgets for joint projects with industry and a nationwide network of "technology extension" centres would help to transfer government research innovations to the private sector.

The people behind these ideas are better defined by their access to Clinton than their ideology. At the top of the list is Gore, whom Clinton has promised will serve as technology czar in his administration, coordinating government science and technology programmes much as Dan Quayle, the current vice president, has tried to reduce industry regulation as head of the White House Council of Competitiveness.

This is a natural role for Gore, whose interest in research issues is so great that some accused him of running for "national scientist", not president, during his failed attempt for the Democratic nomination in 1988. Gore is chairman of the science, technology and space subcommittee of the Senate Commerce, Science and Transportation

Committee and has held dozens of hearings on subjects ranging from global warming and the environment to supercomputers and space. His pet project has been a national high-performance computer network, a modern equivalent to the automotive highway programmes that his father, former Sen. Albert Gore Sr, promoted 30 years ago.

Apart from Gore, those closest to Clinton have little direct experience in science



Schneider coordinates working group.

policy. Magaziner is president of SJS, Inc., a corporate policy consulting company based in Providence, Rhode Island, specializing in international industrial competition. His one notable foray into basic research is not his proudest moment: in April 1989 Magaziner failed to convince Congress to spend \$25 million to create a cold fusion research centre "before the Japanese do". As a Clinton adviser, Magaziner has concentrated on domestic economic policy, including research on 'strategic' technologies.

Schneider, a long-time friend of Clinton, is president of Restructuring Associates Inc., of Washington, DC, another corporate consulting company specializing in competitiveness. He is coordinating a loose-knit science and technology working group which has provided position papers for the campaign. Among those who have contributed to the group's recommendations are Sculley, Young and members of the private Council on Competitiveness (unrelated to the White House office of the same name), including Erich Bloch, the former director of the National Science Foundation (NSF).

It is hard to find a working scientist who has shaped Clinton's thinking. A number of

researchers, including Alexander Rich, a Massachusetts Institute of Technology molecular biologist, have contributed position papers on science issues, but their impact appears slight. The campaign's only statement on basic academic research, released on 12 October, assures scientists that "every dollar we cut from unnecessary defence spending will be redirected back into our university and research community." But it does not specifically mention basic research, and some of Clinton's previous statements — including a sketchy offer to increase research spending by decreasing overhead payments to universities — suggest that he has not spent much time on research issues.

One group hoping to change that is a collection of some 60 researchers, mostly physicists, known as the Council of Scientists and Engineers for Clinton-Gore. The group, organized by Marvin Goldberger, former president of the California Institute of Technology and the Institute for Advanced Studies in Princeton, has yet to meet or issue a policy document, but it plans to focus on science education, the significance of academic research and the need for increased government funding on science. "We're concerned that Clinton is overlooking the importance of basic research", says Goldberger, now at the University of California, Los Angeles.

Although the campaign has so far refused to get into the details of how a Clinton administration would be organized, outside advisers are already working on transition plans. One is William Wells, who until last year was chief of staff for D. Allan Bromley, the current White House science advisor. Wells and others want to make the science advisor a cabinet-level position, equivalent to the president's national security advisor. Another advocate of a stronger science adviser is Richard Bradshaw, a former NSF official who is working with Schneider.

Despite Clinton's outsider image, many of his science policy positions are already Washington staples. The importance of industrially-orientated research, a favourite topic of Walter Massey's, the NSF director, is "very much in line with the Clinton administration's goals", says Bradshaw, and not far from the type of arguments that Bromley made, without success, to his Republican peers. "A lot of our policy [recommendations] come from programmes that were abandoned or underfunded by the Bush administration", he adds. Clinton may portray himself as an agent of change, but his science policy reflects trends that are already well underway. If he wins next month, their pace can only accelerate.

Christopher Anderson

Japan modifies plans for plutonium in wake of protests over shipments

Tokyo. Growing international concern over Japan's plans to ship large amounts of weapons-grade plutonium from Europe has encouraged the head of Japan's most powerful organization for research and development of nuclear power to suggest a new form of nuclear fuel recycling, avoiding the production of pure plutonium that could be diverted to nuclear weapons.

Last week, the president of the Nuclear Fuel Development Corporation (PNC), a semigovernment organization affiliated with the Science and Technology Agency (STA), proposed at a symposium in Tokyo that Japan should consider operating future nuclear power stations on unpurified plutonium, not on pure weapons-grade plutonium derived from reprocessed nuclear fuel (see below). The idea would mark a drastic reorientation of Japan's research and development of nuclear power, but it faces many technical hurdles.

PNC's idea is much more than just a measure to limit proliferation of nuclear weapons. If implemented, it would separate the technology for nuclear fuel recycling from its origins in nuclear weapons.

PNC is probably Japan's richest government-supported research and development organization, receiving about ¥150 billion (US\$1.25 billion) a year from STA, or more than a quarter of the agency's total budget and an additional ¥50 billion from Japan's electric utility companies. The money goes for development of new types of nuclear fuel and nuclear power stations, including the prototype fast breeder reactor Monju, which is expected to go critical early next year after having already cost ¥600 billion.

Monju's appetite for plutonium is driving Japan's plans to start shipping tons of plutonium each year from reprocessing plants in France and Britain. But opponents argue that Japan already produces enough plutonium to meet Monju's needs. A Japanese freighter, the *Akatsuki Maru*, is on its way to

Cherbourg to pick up the first shipment of over one ton of plutonium derived from spent Japanese nuclear fuel reprocessed in France.

The planned shipment has provoked outcry from South Africa to Malaysia to the tiny Republic of Nauru, a 3-km-wide island in the middle of the Pacific Ocean. All are concerned that the ship might be involved in an accident or hijacking on their doorstep.

Officials in STA's Nuclear Fuel Division, trying to distance themselves from the PNC proposal, say that it is simply an "idea" put forward by the semi-government corporation. A PNC official says that STA is awaiting a debate next year by subcommittees of the Atomic Energy Commission on Japan's next five-year plan for development of nuclear power. But PNC is clearly committed to this drastic reappraisal of Japan's plutonium policy.

PNC's idea would require new technology to use crude plutonium in unpurified form mixed with highly radioactive material that is now removed in the reprocessing of nuclear fuel. Present thinking calls for burning of purified plutonium in yet-to-be developed commercial breeder reactors and advanced conventional reactors fuelled with mixed uranium and plutonium fuel. The unpurified fuel would be a less efficient fuel and would require sophisticated robot technology to handle material much more radioactive than pure plutonium and uranium. But it would reduce the amount of high-level radioactive waste from nuclear power plants and could not easily be diverted into the production of nuclear weapons.

Japanese laws require nuclear power to be used only for peaceful purposes. Under its present constitution and international treaties, Japan cannot divert plutonium into production of nuclear weapons. But in less than 20 years the amount of pure plutonium in Japan, including imports from Europe and domestically produced plutonium, will

rise from a few to nearly 100 tons.

Nuclear weapons require the extraction of pure plutonium that, being only weakly radioactive, fits into a light bomb casing. Highly radioactive elements must therefore be removed in the production of plutonium. As a result, the nuclear power industry has not yet tried seriously to develop the technology to handle and manufacture this more radioactive material into fuel.

The PNC proposal is long-term. It has taken Japan 25 years to reach its present level of capability to produce pure plutonium. Development of the proposed new nuclear fuel recycling process would take at least ten to fifteen years and it will not affect facilities already close to operation, such as Monju and the huge Rokkasho nuclear fuel recycling facility under construction in northern Japan.

David Swinbanks

Mail-order notification would replace permits for US field tests

San Francisco. US biotechnology companies will be able to notify the government by mail of field tests of six major crops instead of waiting months for a formal permit under proposed regulations that could be announced this week. The new rules for field tests of genetically engineered organisms would apply to crops modified with specific genes to resist insects or disease or to tolerate stronger doses of certain herbicides.

For months, academics and members of the US Council on Competitiveness, which would prefer even fewer restrictions, have battled with an unusual alliance of environmentalists and the US biotechnology industry, which value continued review, over the contents of the proposed regulations. Earlier this week, the two sides met to settle their differences.

As written by the US Department of Agriculture (USDA), the regulations would exempt from the permit procedure crops and gene modifications already tested under the more than 340 permits issued for nearly 700 US sites over the past five years. Trials of tomatoes, corn, cotton, soybeans, potatoes and tobacco transformed using disarmed *Agrobacterium tumefaciens*, certain virus-coat protein genes, certain virus antisense genes and/or various noncoding regulatory DNA elements, could go forward after a simple notification. All field tests would have to comply with special management techniques, including destruction of the genetically altered plants in the field and moni-

How to handle hot plutonium

Plutonium is extracted from spent nuclear fuel by dissolving fuel rods in nitric acid. Gases given off are collected or safely disposed of in the environment and the aqueous solution is then chemically treated to remove highly radioactive components. Disposal of this high-level radioactive waste remains a significant problem for governments and the nuclear power industry.

Japan's PNC has proposed that these highly radioactive materials need not be removed if robot technology can be developed to handle the material and form it, first, into pellets and then fuel rods. Once made into rods, the fuel could be used by existing nuclear power plants. The new technology would give these plants a second purpose apart from generating electricity: they would become facilities for the burning and reuse of high-level radioactive waste.

D.S.

toring the site afterwards.

The proposed rules are expected to keep a close watch on transgenic plants being developed for the production of therapeutic agents and immunologically reactive proteins. In one proposed field test in Oklahoma, for example, alfalfa plants with part of the cholera toxin gene would be tested for their ability to produce large amounts of biologically active peptide or protein used in the development of a vaccine.

Terry Medley, chief of USDA's biotechnology section, says that not enough is known about the interaction of genetically engineered crops with their environment to remove other regulations, but that the proposed changes should both help to allay public fears about the technology and maintain the competitiveness of new exports.

But some academics call Medley a government bureaucrat who is "bullying" the rules through the system to protect his regulatory turf. Suzanne Huttner, director of the Systemwide Biotechnology Program at the University of California, says the changes do not go far enough and that researchers working on the six crops should be permitted to notify US agriculture officials the day they plant. She would also like to exempt researchers from the need for a permit on new plants or other organisms not previously identified as plant pests.

In a letter to the White House Council on Competitiveness, Huttner and other university researchers asked for the removal of all special rules for biotechnology field tests. Huttner says that USDA rules add to the cost of research and to public fears of the technology by focusing on the process of genetic engineering. The proposal "definitely sells out basic research", Huttner says. "It guarantees that these products will sit on laboratory shelves and never see the light of day."

But environmentalists and industry, usually on opposite sides of regulatory questions, have joined to support the USDA proposal. The Industrial Biotechnology Association believes that continued federal review will help it to win public acceptance and stave off more restrictive state actions. North Carolina, Minnesota and West Virginia, for example, have enacted their own rules covering environmental release of genetically engineered organisms. "We believe [the USDA proposal] is a sound step and scientifically justified," says Alan Goldhammer, director of technical affairs for the US trade group. US companies should be able to notify USDA for most of their field tests, he predicts.

At the same time, environmental groups are now concerned that USDA may abandon its cautious approach to regulating genetically engineered plants and choose same-day notification, which would prevent environmentalists or states from commenting on field tests before they begin or replace a detailed list of exemptions with a broader rule.

Sally Lehrman

Nobel Prizes go to Caltech chemist, CERN physicist

Physics prize

London. This year's Nobel Prize in physics has been awarded to a man whose inventions have formed the basis of high-energy particle detection for the past twenty years. Georges Charpak, 68, who has spent most of his career at CERN, the European Laboratory for Particle Physics, is cited in particular for the invention, in 1968, of the multiwire proportional chamber — a device whose

descendants are active in most high-energy physics experiments operating today.

From the earliest days of nuclear physics, the identification and study of subatomic particles has relied on ways of making visible the trail of ionization created by an energetic particle as it travels through matter. Wilson's cloud chamber, Powell's high-quality photographic emulsions and Glaser's bubble chamber — all Nobel-prize-winning inventions — revealed a host of new particles from cosmic rays and accelerators, starting with the discovery of the positron in 1932.

But as physicists started to look for rarer particles, they needed an automated way to find exotic tracks among millions of more common events. They also needed speed: at only about one 'snapshot' per second, the bubble chamber could not keep up with increasingly intense accelerator beams.

Charpak's contribution was to see how to combine high spatial resolution with the speed and output offered by existing electronic detection methods. He returned to a device that had been used as early as 1908 by Rutherford and Geiger: the proportional counter, a gas-filled tube with a thin, high-voltage wire at its centre. The passage of an energetic particle ionizes the gas, triggering a pulse of electrons onto the wire. Charpak realized that an array of such wires, spaced 1-2 millimetres apart in a gas-filled chamber, would act as independent sensors and could thus be used to track a particle's path. The electrical pulses could be sent directly to a computer, with a detection rate as high as a million particles per second per wire.

In their 1968 paper, Charpak *et al.* also proposed an important refinement of the technique: measurement of the 'drift' time between the initial ionization of the gas and the arrival of the electron pulse at the wire, to provide even better spatial resolution. But

Charpak has lately turned away from particle physics to apply his detectors to biology and medicine. His spherical drift chamber is a highly efficient detector for the X-ray determination of protein structures, and his imaging proportional chambers bring far greater sensitivity than photographic film to the autoradiography of biological samples.

Laura Garwin

Chemistry prize

Pasadena. A theoretical chemist at the California Institute of Technology has won the 1992 Nobel Prize in chemistry for his contributions to the theory of electron-transfer reactions. Rudolph Marcus, 69, received the prize for a series of papers, published between 1956 and 1965, that explained puzzling variations in reaction rates in redox processes and made counter-intuitive predictions not verified experimentally until the 1980s.

One of Marcus's first contributions was to elucidate the role of surrounding solvent

molecules in determining the rate of redox reactions — reactions that are fundamental in energy storage, transport and conversion. Marcus further determined that the relationship between the driving force of an electron-transfer reaction and the reaction's rate is described by a parabola. This im-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Rudolph Marcus

plies that as more driving force is applied to a reaction, its rate at first increases but then begins to decrease in the so-called inverted region. Chemists greeted this surprising insight with considerable scepticism until its experimental verification in 1985.

Although the royal Swedish Academy of Sciences cited only Marcus's electron transfer work, he has made contributions in several other areas. For example, his name is the M in RRKM theory (the others are Rice, Ramsperger and Kassel), an attempt to explain unimolecular reactions that has found wide application in chemical physics. He has also done important work in transition-state theory and in the theory of collisions and bound states. Marcus has continued his research in electron transfer, more recently turning his attention from simple reactions to reactions in complex molecules such as the proteins involved in photosynthesis.

Robert Finn

Talking telephone numbers

SIR — Your comments on telephone numbers (*Nature* 358, 440; 1992) oversimplify the situation.

(1) The currently normal 10-digit string we dial is actually a 9-digit identifier preceded by a single-digit access code. There is a trade-off in the time taken to use an access code: if it were not used to select out-of-area calls, it would be necessary to dial 9 digits even for the local calls for which we normally use 6 or 7 digits.

(2) Historically the leading digits had geographical significance, but their importance now is that they carry charging information. Until the day arrives that all standard point-to-point calls within the United Kingdom are charged at the same rate, it will be necessary to allocate number ranges to geographical areas. In addition, other number ranges (800, 345, 500, 898 and others) attract specific charge rates independent of destination. The cost to business of not being able to monitor telephone bills by checking the codes dialled could well outweigh the cost of using an extra digit.

(3) In the longer run, the extra digit can be regarded as a selector of both carrier (British Telecom, Mercury or other) and service (for example mobile telephones, personal location-independent numbers) and so it may become the primary indicator of charge rate.

(4) Already in parts of the United States there are more telephone numbers than people so the 10^8 limit you suggest for the United Kingdom is likely to be exceeded early in the next century.

To summarize, the main factor in the inefficient use of the numbering range is not the need for geographical information as such, nor the historical limitations imposed by old technology, but the unwillingness of the consumer (and Oftel, the regulatory body) to accept that with current technology almost all the cost of a call arises in the local loop, and so is independent of the distance to the called party.

John Littler

*Telecom Potential,
Clevedon Hall, Victoria Road,
Clevedon, Avon BS21 7RQ, UK*

SIR — As you point out, modern electronic switching works equally well with any sequence of digits, and telephone companies would have more numbers available to them if they didn't have to reserve specific sequences for specific areas.

What you did not consider, however, was who is on the other end of all that modern electronic switching. Even in this era of faxes and modems, most telephone calls are still placed by hu-

mans, who have to try to remember an ever increasing number of telephone numbers. Humans, unlike computers, remember numbers better when the numbers are in groups and when the groups have some sort of reference. Thus, 912027372355 (the telephone number of *Nature's* Washington office) is almost impossible to remember unless we take it in groups: 9 (that gets me an outside line), 1 (long distance), 202 (the Washington DC area code), 737 (the exchange), 2355 (the number). To remember this number, all I have really to remember is Washington, the exchange, and the number. Area codes and exchanges I call frequently are easy to associate with specific geographical areas and thus easier to remember than random sequences. The scheme you suggest would make it much harder on us poor humans.

Michael Blecher

*AAPG Publications,
PO Box 979,
Tulsa, Oklahoma 74101-0979, USA*

Consciousness

SIR — The discussion of consciousness¹⁻³ suggests the lack of a consistent definition. Consciousness is clearly a result of neuronal activity, because it is suspended by sleep, mechanical trauma and drugs such as anaesthetics. As we humans evolved from other mammals, our brains are basically the same as theirs, so that the question also arises of whether consciousness differs qualitatively or merely quantitatively from the brain activity of apes or dogs. Unfortunately, they cannot tell us whether they are conscious. But this highlights my conception of consciousness, the ability to talk silently to myself.

Language is unique to humans, and all our civilization depends on it. Teachers know that full understanding of a subject often comes only when it has to be put into words in order to communicate that understanding to students. We are capable of transmitting and receiving complex information encoded in abstract acoustic or visual signals (talking and reading). Suppose that an individual's brain encodes the information it is 'aware of' into language, and then interprets the stream of chosen words much as if it came from an external source, checking the accuracy of the description, and possibly generating corrections. That activity would seem to fulfil the basics of consciousness.

Let me suggest a model: individual A keeps up a running commentary on his actions and perceptions to individual B,

who listens, interprets and, when he feels like it, replies with observations and suggestions that A uses to modify his actions and perceptions. B is effectively the consciousness of A. If A and B are the same person, we have a conscious person. Within one brain, the communication could include neuronal representations of phenomena as well as abstract language, which would be available to other animals. But only humans have the subtlety of language to fine-tune the process. This does not solve the problem of consciousness but makes human consciousness a consequence of the development of language, as Jaynes⁴ has argued.

Hugh L. Fletcher

*Queens University,
Division of Genetic Engineering,
97 Liston Road, Belfast BT9 7BL,
Northern Ireland, UK*

1. Gray, J. *Nature* 358, 277 (1992).
2. Josephson, B. D. et al. *Nature* 358, 618 (1992).
3. Sacks, O. *Nature* 358, 618 (1992).
4. Jaynes, J. *The Origin of Consciousness and the Breakdown of the Bicameral Mind* (Houghton Mifflin, Boston, 1990).

True or false?

SIR — Writing about my book *The Facts of Life: Shattering the Myth of Darwinism*, the author of your leading article (*Nature* 358, 698; 1992) says that I believe evolution to be false, that natural selection is a 'pack of lies' and that I believe the Earth to be young. Any objective reader of my book will find that none of these statements is true. I accept evolution; but I present evidence that the neo-Darwinian mechanism cannot wholly account for it. I do not think natural selection a pack of lies; I present evidence that it is incapable of leading to trans-specific evolution. And I do not claim the Earth is young, but present evidence that currently accepted methods of geochronometry are seriously flawed and should be re-examined.

Instead of addressing the scientific issues raised by this evidence, you dismiss it simply by placing quotation marks around the word evidence, implying that what I have written is not really evidence.

I am a professional science reporter and if I have misrepresented any of the scientific work referred to I shall be glad to receive specific details.

Richard Milton

*Beech Hurst,
10 Pembury Road,
Tonbridge, Kent, UK*

□ So why does the book jacket proclaim that the book contains "[t]he evidence against the Darwinian idea that chance is the mechanism of evolution"? — Editor, *Nature*.

Struggles to correct published errors

Fredric M. Menger and Albert Haim

The following is an account of what happened when two researchers independently tried to correct errors in articles published in the leading US chemical journal.

It has recently been said of the US scientific establishment: "As the number of scientists reaches 1 million and their share of the nation's federal budget reaches \$25 billion, the demands for greater accountability and openness are understandably more insistent"¹. This increasing need for openness prompts us to report our attempts to disclose errors in the leading journal of the American Chemical Society. We show that the peer-review system broke down when papers with egregious errors were accepted for publication. Moreover, we demonstrate that the self-rectifying mechanism for maintaining the integrity of science initially failed when the editors refused to publish corrections in the journal in which the errors originated. Our experiences are revealed here because they might represent problems that have become ingrained within the chemistry establishment. Solutions to the problems are possible only if the scientific community is made aware of what can happen.

Events began with the publication of two articles^{2,3} in the *Journal of the American Chemical Society* (henceforth *JACS*) by Professor Ronald Breslow of Columbia University. E. Anslyn was the coauthor of one paper² and D. Huang was the coauthor of the other³. Both papers dealt with the kinetics of cleavage and isomerization of dinucleotides (reactions important as ribonuclease models). To document the course of events after the appearance of these papers, we will use quotations from letters, reports and articles taken from our 'Breslow file', now several inches thick.

It was immediately apparent to us (working independently and unaware of each other's efforts) that both papers were beset with problems. Rate-constant units were garbled; experimental error was very large; pH and ionic strength had not been properly controlled; straight lines were drawn through graphs having only two experimental points although equations predicted non-linear plots; and rate constants, invoked to fit the data, never had their physical significance explained. Most puzzling of all, however, was the reporting of negative rate constants. Even the title of Breslow and Huang³ stressed this aspect of the work: "A Negative Catalytic Term Requires a Common Intermediate in the Imidazole Buffer Catalyzed Cleavage

and Rearrangement of Ribodinucleotides". Usually, of course, rate constants are positive numbers.

On 8 March 1991, one of us (F.M.M.) decided to alert the readers of *JACS* to these difficulties, and submitted a short note to the senior editor, Professor Allen J. Bard. Bard asked one of his associate editors, Professor Richard L. Schowen, to handle the manuscript. Two weeks later, the manuscript was returned without the benefit of reviews by independent referees. Schowen wrote: "I don't think this paper is any good. Breslow and Huang never said they measured a negative rate". It was difficult to understand the reply as Breslow and Huang had published an entire list of negative rate constants.

Independently, A.H. submitted a critique of Breslow's papers to Bard on 19 March 1991, pointing out that Breslow's proposed mechanism was incompatible with his own kinetic measurements. Moreover, if one inserted the Anslyn and Breslow² rate constants into their published equations, many of the published plots could not be duplicated. Bard also assigned this manuscript to Schowen, who returned it without seeking referees' advice. Schowen wrote: "It may well be that there are flaws in the Breslow papers. In order to publish criticisms of the papers in *JACS*, however, it would be necessary to present material which is itself a contribution to the understanding of these systems as opposed to merely correcting, or as it seems to me that you are doing, taking issue with Breslow's interpretations."

Following the rejection of his paper, F. M. M. felt obliged to contact Bard to voice his unhappiness over Schowen's decision. In the meantime, A.H. wrote to Schowen (and sent a copy to Bard) on 24 April 1991: "The peer review system has broken down and you are not correcting the problem. How is science to correct itself if *JACS* does not allow the publication of articles that point out errors?" Bard decided that both manuscripts merited formal reviews by independent referees and asked Schowen to proceed accordingly. Thus, on 28 April 1991, F.M.M. submitted a manuscript entitled "The Negative Rate Constants of Breslow and Huang". A.H.'s paper, submitted on 7 May, 1991, was entitled "Mechanisms, Kinetic Models, Functional Dependencies, and Negative Catalytic

Terms in the Imidazole Catalyzed Cleavage and Isomerization Reactions of Dinucleotides".

We must now backtrack slightly. Before submitting his paper, A.H. had written to Breslow about the problems he discovered in Anslyn and Breslow². Breslow passed this letter to Anslyn, who by then was at the University of Texas. On 8 April 1991, Anslyn wrote Breslow a letter dealing with these questions (a letter that was passed on *in toto* to A.H.). From the standpoint of F.M.M.'s paper, the most interesting sentence of Anslyn's letter was the following: "The fact that 7B has a negative rate constant intercept is an artifact of our data correction process". Thus, Breslow's own collaborator had alerted him, in writing, to the fact that the negative rate constants were suspect.

Both our manuscripts were rejected by Schowen in July 1991. The referees' reports turned out to be a strange mixture. On the one hand, they were critical of the Breslow papers as the following quotes show:

There are undoubtedly problems with Breslow and Huang. The reactions are very slow and cannot be followed directly. Most of the so-called buffer runs are not buffered; ionic strength was not kept constant; the extrapolations described are optimistic; and the presentation is vigorous but simplistic. Many people — including one of my research students — have pointed out these shortcomings.

The rate constants given by Anslyn and Breslow do not reproduce the lines given in their paper for isomerization.

On the other hand, both referees ultimately concluded that our papers should be rejected.

One must recommend against publication although it is to be hoped that Huang and Breslow will publish a full paper which gives a corrected and intelligible account of what was done.

F.M.M. appealed to Bard to overrule Schowen's decision. On 2 August 1991, Bard wrote to F.M.M.: "Based on my reading of this material, I feel that Professor Schowen's decision not to publish your paper was the correct one." He continued: "I am less familiar with the representation of rate constants as negative numbers, but I am not sure this is improper." Thus F.M.M. had apparently been unable to persuade the editors that

the negative rate constants, on which Breslow and Huang³ was based, were not simply the result of a scholarly formalism, a semantic flourish or graphs with negative slopes. Instead, negative rate constants arose by improperly adjusting raw experimental data for a large background rate.

F.M.M. wrote to Bard by return, stating "I also recognize that errors are part of science and an inevitable component of scientific articles. One cannot turn journals into a debating ground for minor mistakes and oversights. If the Breslow papers had contained only minor peripheral errors, I would not have written a critique. But the papers were so laden with incorrect equations, statements, and conclusions that I (along with Dr Haim) felt a duty to respond. In my opinion, the Journal was not well served by allowing this bad science to remain uncorrected within its covers."

Following rejection of his paper, A.H. sent a fax message to Schowen (sending a copy to Bard) pointing out the American Chemical Society ethical guideline: "If an editor is presented with convincing evidence that the main substance or conclusions of a report published in an editor's journal are erroneous, the editor should facilitate publication of an appropriate report pointing out the error and, if possible, correcting it. The report may be written by the person who discovered the error, or by an original author". A.H. then wrote to Schowen (copying the letter to Bard), requesting publication of the manuscript (or a suitable revised version) on the basis of the ACS ethical guidelines. Bard rejected the request on 2 August 1991, to which A.H. responded on 11 August 1991:

It is disturbing to reflect upon the fact that the editors had an opportunity to rectify this unsatisfactory situation by informing the scientific community of the grave errors in Breslow's papers but refused to do so. Indeed, this whole affair represents a dark chapter in the annals of Science.

On 18 September 1991, Breslow unexpectedly wrote to A.H., stating: "I don't understand what you are so excited about". "Are you being led astray by a notoriously unstable individual?" There was no doubt to whom Breslow was referring.

At this point, we had both failed in our effort to correct errors in the literature. Unwilling to let the matter drop, A.H. submitted his paper to the *Journal of Physical Chemistry*. A lone referee wrote the following sentences (abstracted from a long report):

To this reviewer the greatest mistake that Breslow and co-workers did was to subtract out the non-buffer catalyzed reaction before

attempting to fit their data to a particular mechanistic scheme. This led to them reporting ridiculous negative rate constants in the Breslow and Huang paper.

Haim is right that there are serious problems with the Breslow papers, but I am not sure his contribution is going to help much. This reviewer is not convinced that the general outlines of the proposed mechanism are wrong; it is just that Breslow and co-workers have not proven (or disproven) it.

A.H.'s paper was rejected.

In the meantime, F.M.M. had sent his manuscript, still entitled "The Negative Rate Constants of Breslow and Huang", to Professor Clayton Heathcock, senior editor of the *Journal of Organic Chemistry* (JOC). Heathcock asked one of his associate editors, Professor Andrew Streitwieser, to handle the manuscript. On 16 August 1991, Streitwieser accepted it. The report from the sole referee is quoted in its entirety:

I recommend the publication by Fred Menger, and I am very surprised that JACS would not accept the contents of this Communication in some form. On first reading the Breslow, Huang manuscript I was dumbfounded by the three figures, each with a linear plot of two points, and the negative rate constants. Clearly, there is no way that the rate expression provided by Breslow could provide negative rate constants. I sat down with [name withheld] and he agreed with me that a horrible mistake had been made. Subsequently, others in other portions of the country, who need not be named, agreed that there was something terribly wrong with this study. I myself could not understand how the Communication could have gotten through the referee process and have passed an editor. I believe Fred Menger in putting together this Communication has done what had to be done.

F.M.M. could not resist sending both Schowen and Bard a copy of this report. Their reactions were quite different but equally perplexing. Bard wrote: "I really am pleased that your Communication on the Breslow/Huang paper was accepted by JOC." Schowen wrote: "Nothing you sent JACS enumerated aerated any real errors, horrible or otherwise."

F.M.M. received many letters after publication of the JOC paper⁴, including an unsolicited letter written to Bard by a Nobel laureate:

It has come to my attention that the enclosed comment, published by the *Journal of Organic Chemistry*, was turned down by the *Journal of the American Chemical Society*. This I find extremely disturbing. It seems to me there is a long-established scientific etiquette which says that papers pointing out errors should be published in the same journal in which the original paper appeared.

It is now necessary to address the issue of "footnote 6" in the JOC article. The footnote read:

Dr Albert Haim of Stony Brook simultaneously and independently uncovered a variety of other problems with the Breslow manuscripts (e.g. the reported rate constants and equations do not fit the theoretical plots). His analysis will be published elsewhere. Neither of us was permitted to publish our work in the journal where the errors originated.

On 1 November 1991, F.M.M. received a letter from Heathcock stating that an erratum would be published in the earliest possible issue of JOC to state that the above passage was "inserted during the galley proofs and did not appear in the version of the manuscript that was approved by the editor; the editor would not have permitted publication of the manuscript with this passage"⁵.

F.M.M. had indeed altered footnote 6 at the proof stage. The original, approved manuscript had read:

Dr. Albert Haim of Stony Brook simultaneously and independently uncovered a variety of other problems with the Breslow manuscripts. His analysis will be published elsewhere. Attempts to document these errors in the journal from which they originated were unsuccessful.

Admittedly, the recasting of the sentences in proof rendered them more acerbic. Although footnote 6 is a correct statement, and can be proved to be so, and although the footnote was a modification and not an "insertion", the change in proof was an error of judgment. Any modification should have received prior approval from the editors. F.M.M. sent an apology to Heathcock and, via this account, extends it publicly.

On 2 December 1991, F.M.M. received another letter from Heathcock with a shocking paragraph: "We have received requests that your paper be retracted. One of these comes, of course, from Breslow, but we have obtained a totally independent request as well". There was, of course, no indication as to the identity of the second critic, a point to which we shall return later. Heathcock ended: "These are significant charges that require responses on your part. Are these statements correct? Can you indicate to me why your paper should not be retracted?"

F.M.M. responded immediately to the critic's statements, and on 27 December 1992, Streitwieser wrote to F. M. M.: "I consider your responses to be satisfactory and no retraction will be made". He continued, "you may be interested to know that I went through a careful kinetic analysis myself of Breslow's mechanism. The problem with having k_w (i.e. background) terms in both the numerator and the denominator is that a dissection into a water part and a buffer part is no longer possible". "This result

has been communicated to Professor Breslow".

It is now necessary to take a short detour. Another researcher, unknown personally to either of us, exchanged correspondence with Breslow on another matter. Breslow wrote a letter dated 8 January 1992 in which he states: "A *JACS* editor who is an expert kineticist has written demanding that the Menger paper be retracted because it is fraudulent". This remark provides telling information as to the identity of the person behind the "significant charges" sent to Heathcock at *JOC*. It is unclear how Breslow learned of the editor's action or whether the editor was aware that Breslow was using knowledge of the retraction demand for his own purposes.

Again we must backtrack. On 10 September 1991, F.M.M. wrote to Professor Ernest Eliel, president of the American Chemical Society, explaining the problems with *JACS*. Three days later, Eliel referred the complaint to Professor Jean'ne Shreeve, chairman of the ACS committee on publications. F.M.M. then sent Shreeve all the relevant papers as they appeared, so that she had in her possession the main part of our Breslow file.

By 8 March 1992, F.M.M. had not heard from Eliel or Shreeve, so he wrote to Eliel inquiring as to the status of the ACS review. Eliel replied, in a letter dated 20 March 1992, that Shreeve was no longer chairman of the publications committee. The file had never been transferred to her successor. All attempts to contact Shreeve, including a certified letter, failed to elicit a response. In his letter, Eliel said "I continue to think that the cause of science has been adequately served." "I feel ACS's hands and my own are clean".

In a letter to A.H. dated 23 October 1991, Schowen had expressed the opinion that Breslow's papers "represent a very good contribution to defining the baseline behaviour of dinucleotides". In another letter dated 13 November 1991, Schowen wrote to A.H.: "My opinion of your criticism is that it is inappropriate". Thus, when A.H. submitted a new version of his paper to Bard on 18 November 1991, he requested that an editor other than Schowen deal with the manuscript. Bard refused, but he did ask A.H. for a list of ten suitable chemists from which Schowen could select referees. On 15 January 1992, Schowen wrote to A.H., saying that he had obtained four reviews. One of these, from Breslow himself, was negative:

If Haim tries to publish it elsewhere, in spite of its flaws, he is either unable to grasp simple arguments or dishonest enough to publish what he knows is wrong. He should not imitate another fraud perpetrator in this.

Another referee wrote the following:

It is deplorable that Breslow's papers on these hydrolyses were ever published. It is immediately obvious that Anslyn and Breslow is sloppy: the ambiguity over the state of protonation, the use of "buffers" of zero or infinite buffer ratio, the failure to control changes in pH and ionic strength (which probably would not even compensate a specific ion effect at 2M buffer), lines in Figs. 1B and 2B (which duplicate 1A and 2A and should not have been included) that erroneously miss the origin, confusion of rates and rate constants. How could such manuscripts pass the referees?

Despite the above, the referee did not consider Haim's paper a sufficient advance in understanding to justify publication. "The proper remedy is to do the experiments correctly".

A third referee voiced the opinion that the matter was best dropped. He or she wrote:

It could lead to a reputation, rightly or wrongly, of the author being a nitpicker and Breslow would certainly fight back loudly. Who needs such things?

The final referee, an authority in modern bio-organic chemistry, sent a signed review stating that "I do not believe that it is proper to delay any longer and recommend that Haim's paper be published in *JACS*". It had been 11 months since A.H. first submitted a manuscript for publication.

Schowen did not take the advice of the fourth referee, but suggested A.H. revise the manuscript without any commitment to publication. On 15 March 1992, A.H. submitted a revised manuscript, and in his cover letter he asked: "Will you, as one of the gatekeepers of Science, knowingly perpetuate the fatal flaws in Breslow's papers within the pages of *JACS*, or instead exert your responsibility and duty as associate editor and help my extraordinary efforts to expose major flaws in Breslow's work?"

On 12 May 1992, Schowen rejected A.H.'s manuscript. Schowen's letter of rejection and the accompanying referee's report are too long to quote in their entirety. It is fair to state, however, that no referee ever defended Breslow's work as being correct. There were, however, differences of opinions as to how to deal with the problem. For example, one referee recommended publication:

I can sympathize with Haim's attitude that *JACS* is the appropriate medium to call attention to Breslow's errors which are truly gross.

Another referee did not:

The present paper would simply call attention to Breslow and his erroneous details rather

than deal with the old, and clear, literature which is to be read and enjoyed.

In his letter of rejection to A.H., Schowen wrote: "I don't think anyone has ever suggested that a paper pointing out serious errors in another publication should not be published. As far as I know, everyone thinks such a paper should be published". Yet Schowen's own referee had called Breslow's errors "truly gross". A.H. again appealed to Bard, and on 12 June Bard accepted the manuscript for publication as a full paper in *JACS* (ref. 6).

As the dust settles, it is comforting to reflect that the system ultimately worked. After all, both of us succeeded in getting our papers published. Yet this was accomplished only at the cost of considerable anguish to both of us. Few people, we presume, would be willing to go through this experience. Certainly, had either of us known at the beginning that the 'Breslow file' was to become inches thick, filled with frustration and insult, we would have been reluctant to pursue publication in the first place.

Two problems are involved. The first is the mishandling of the original publications, which many people have come to regard as substandard. The second is the position taken by the associate editor after flaws were pointed out. That position can be described only as defensive and evasive. Without attributing motivation for his actions, we simply state that we believe them to have been inimical to the best interests of science.

Editors, perhaps more than anyone, suffer from an almost unmanageable flood of literature. The burden is passed on to referees who, at times, are too busy to read manuscripts carefully. In this manner, errors creep into journals. Certainly one of the best protections is for editors to permit scholarly rebuttals when a published article is read carefully and when serious errors are thereby uncovered. The very knowledge that such a policy is in force will lead to the best policing of all — from the authors themselves. □

Fredric M. Menger is in the Department of Chemistry, Emory University, Atlanta, Georgia 30322; Albert Haim is in the Department of Chemistry, State University of New York, Stony Brook, New York 11794, USA.

1. Hiltz, P. J. *The New Republic* 24 (18 May 1991).
2. Anslyn, E. & Breslow, R. *J. Am. chem. Soc.* **111**, 4473–4482 (1989).
3. Breslow, R. & Huang, D. *J. Am. chem. Soc.* **112**, 9621–9623 (1990).
4. Menger, F. M. *J. org. Chem.* **56**, 6251–6252 (1991).
5. *J. org. Chem.* **56**, 6960 (1992).
6. Haim, A. *J. Am. chem. Soc.* (in the press).

ACKNOWLEDGEMENT. We thank George Hammond for assistance.

The model for almost all seasons?

The solution by Onsager, almost exactly half a century ago, of a simple-minded two-dimensional model of a ferromagnet has been a powerful stimulus in theoretical physics.

A MODEL is an approximate description of reality, right? And a good model is a model which, on the one hand, is comprehensible in the sense of providing an image for the mind and which, at the same time, is calculable. Everybody agrees.

The most familiar example is the model of the perfect gas as an assemblage of rigid elastic spheres, billiard balls in textbook language. Once Maxwell had shown that a collection of such spheres would indeed obey the rules required of a perfect gas, the rigid-sphere model of kinetic theory was destined for a virtually infinite life. Much of what has since been called physical chemistry consists of the definition of the ways in which atoms and molecules cannot be accurately represented by rigid elastic spheres, and of the offering of explanations.

In retrospect, it would have been good if Maxwell and his contemporaries had been a little more adventurous. The best part of half a century went by before people started looking seriously at the properties of the rigid-sphere liquid, for example. But for the purposes of the kinetic theory, and once the supposition that atoms are rigid elastic spheres has been shown to be a good first approximation, why not have gone a step or so further, and have tested the model to destruction?

Over the years, that has inevitably been done. Relaxing the condition that the collision between atoms should be elastic, or that momentum should be conserved, leads straightforwardly to predictions of the departure of real billiard-ball gases from the laws of the perfect gas and to the calculation of the virial coefficients in terms of the parameters of inelasticity, whatever they may have meant physically. Similarly, it would have been good to know what the late nineteenth century made of the non-spherical, perhaps ellipsoid, model of the atom. There may still be essays in this direction by disappointed authors still to be rescued from late-Victorian attic rooms.

It is difficult to think of a more robust (in the sense of a more generally applicable) model of the real world unless it is the tautologous continuum model of three-dimensional reality. Formally at least, you can calculate almost anything that way. The deformation of a solid object under the influence of an external field of force? No problem.

Simply imagine that the body is made up of a collection of parallelepipeds cemented together by internal forces, calculate the deformation of each of them in terms of the

coefficients of the supposed elasticity tensor, put all of them together and require that the forces across each face cancel out. The outcome is a differential equation of some equivalent thereof. But a little reflection will show that these elegant ways of talking, which brought great fame to late-Victorian Cambridge, are not much more than ways of doing calculus in three dimensions (whence the term "tautologous"). They have little to say about physics.

If the three-dimensional continuum is an arid venue to explore, there is luckily a class of models that compare well, in utility and versatility, with the simple hard elastic sphere as a model of an atom in a perfect gas: the Ising model, as this potential paragon among physical models is known. Ising, for what it is worth, was an Austrian who, in the early 1920s, believed it would serve the public interest if there were a simple and stylized model of a ferromagnet that would be mathematically tractable.

So why not take, to begin with at least, the simplest possible case, that of a three dimensional cubic lattice, and site a little magnet at each vertex? Further simplifications immediately suggest themselves. If the little magnets represent real dipole magnets, with fields of force declining as the third power of the distance, surely it will suffice to count only nearest neighbour interactions. And then, with the arrival of quantum mechanics and the doctrine of electron spin, surely it will make sense to suppose that the little magnets can point in only one direction (that of the external field) or the opposite?

So there emerges the Ising model of a quantised ferromagnet: little dipole magnets at the vertices of a regular lattice, and able to point in one direction or the other. There need not be an external magnetic field, but, on the other hand, there may be. The problem is to calculate the degree of magnetic ordering within the three-dimensional array. It sounds simple. As it happens, the problem is insoluble. For the best part of half a century, people have broken their heads on it, but have failed to solve it. Worse, there seems to be no proof that the conundrum cannot be solved analytically, which means that people keep on breaking their heads.

So how can such a model serve any useful purpose? The simple answer is that the model is at least an easy means of defining problems and of representing them mathematically. You might say, for example, that the energy of interaction between two neighbouring sites in the lattice is either some quantity J or $-J$, depending on whether the

little magnets are pointing in opposite or the same directions. To make the ferromagnetic model into that for an anti-ferromagnetic, simply make J negative. And so on. So what a pity that the model is incalculable.

That is how it appeared to have remained until the early 1940s. By then, it had been recognised that the Ising model would also be a splendid model of binary alloys such as β -brass, whose practical value rests on a marked transition from an ordered (and tough) state to one in which the lattice atoms are randomly arranged, and which is a malleable metal by comparison. And if only the model could be solved, would it not be possible to make a model of a liquid in equilibrium with its vapour by letting occupied lattice sites represent real atoms, interacting with their neighbouring sites if they are also occupied, but supposing that empty sites represent mere vacuum?

Only in the early 1940s did the late Lars Onsager provide a half solution, a way of calculating the properties of a two-dimensional lattice. Although this calculation was exact, on the face of things it had little to do with real ferro-magnetism; most conspicuously, it did not uncover the finite change of entropy (signalled by the emission of finite amounts of heat on cooling) of real ferromagnets as they are cooled through their transition or Curie temperatures, but only an abrupt change of the specific heat. But that, of course, is how two-dimensional ferromagnets probably behave.

It is quite remarkable what Onsager's demonstration has since accomplished. Almost trivially, given that the energy of a particular configuration of spins on a two-dimensional lattice is a measure of the number of nearest-neighbour sites which are occupied, the technique becomes a way of dealing with random lattice walks. By supposing that interconnections consist of chemical bonds, it is a way of calculating the configurations of polymer molecules.

Ambitions to generalise from two to several possible spin-configurations at every lattice point have been successfully accommodated, while the arrival of high-speed computers has made possible numerical approximations to the solution of the three-dimensional problem. Add in the remarkable way in which similar calculations are now routinely applied to the solution of problems in Quantum Chromodynamics, and you have the basis for a claim that the simple-minded Ising lattice is the most versatile model of them all.

John Maddox

Carbon onions introduce new flavour to fullerene studies

H. W. Kroto

THE notion that graphite, composed of flat sheets of carbon hexagons, is the most stable form of carbon is called into question on page 707 of this issue¹. D. Ugarte has found that soot, when annealed by intense electron irradiation, naturally transforms itself into giant nested shells of carbon, possibly the onion-like relatives of fullerene.

Through synthesizing large quantities of C_{60} fullerene, Krätschmer *et al.*², in 1990, finally confirmed that this all-carbon molecule is indeed a spheroidal shell. Our proposal³, in 1985, that the stable C_{60} cluster detected during laser ablation of graphite has this form was initially greeted, by some, with natural scepticism. After all, carbon, it was generally held, invariably forms flat sheets of atoms in hexagonal arrays, stacked as graphite — deemed the most stable form of the element. If, however, the C_{60} species, which had been created spontaneously in the chaos of a hot carbon plasma, was in fact a closed spheroidal cage, then the whole concept of the intrinsic flatness of graphite might need some re-evaluation.

It soon became apparent that on the scale of a few tens to a few hundreds of

carbon atoms, graphite-like sheets of hexagonally arrayed carbon atoms would be unstable because of dangling bonds at the edges; this instability could be relieved by closure into a spheroidal network. Perhaps in bulk these edge instabilities would be insignificant, but on a small scale the need to satisfy the dangling bonds is probably decisive and leads to energy-driven closure. Closure could be achieved with such flat (hexagonal) sheets by the inclusion of 12 pentagons, as David Jones⁴ pointed out, under his *nom de plume* Daedalus, in 1966. In fact, a close look at graphitic material shows that the effectively infinite flat sheets of graphite, which are presumed to form spontaneously, are actually rather elusive — to the point of non-existence.

After the original fullerene proposal was made, Iijima's beautiful transmission electron microscope studies of concentric graphitic shell structures⁵ became a focus of interest, as they appeared to bear some relation not only to the structure of C_{60} but also to the mode of its formation. K. G. McKay and I suggested⁶ that Iijima's structures might be explained as quasi-icosahedral concen-

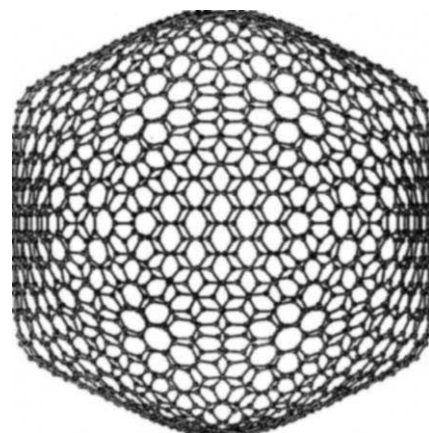


FIG. 1 Predicted structure of the giant fullerene C_{1500} , which would form the fifth shell of Ugarte or Iijima's carbon onion.

tric shells that grew through a spiralling-network mechanism⁷. These are essentially a set of concentric giant fullerene shells (Fig. 1) one inside the other, and electron-microscope simulations (Fig. 2) add significant support for this proposal. Although it was originally thought possible to produce a set of perfectly closed onion-like shells, it also seemed likely that this would occur rather infrequently. The results of Smalley and colleagues⁸, however, indicate otherwise, and closure might be more the rule than the exception — at least during gas-phase nucleation.

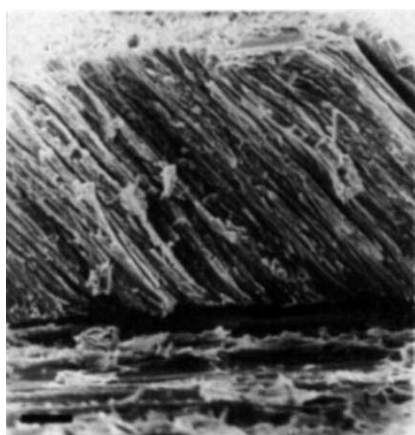
Now the elegant electron-microscope study of Ugarte indicates that closure may also occur readily in the condensed phase. By high-temperature processing of small carbon particles, Ugarte has produced kernels which exhibit a remarkable degree of spheroidal structure. He finds that under the microscope's electron beam, small fragments of material produced in a fullerene generator can self-assemble or rearrange themselves into concentric spheroidal graphitic shells very similar to the structures observed by Iijima⁵. But Iijima believes his particles were produced in a different way, perhaps by nucleation in the gas phase (personal communication).

The C_{60} buckminsterfullerene and other results⁸ indicate that 60–600-atom aggregates of carbon atoms must be more stable as a closed spheroidal cage than as a flat sheet. Now, Ugarte's results show that carbon atom aggregates of 10^6 – 10^7 atoms appear to be more stable as onion-like structures than as flat sheets. (One can count the atoms because, assuming the object is perfect, the central shell has 60 atoms, the next 240, the next 540 and so on; Fig. 2 of Ugarte's article reveals 60 or more shells.)

Ugarte's results are also interesting for the light they cast on the mechanism by

NATURE · VOL 359 · 22 OCTOBER 1992

BESIDES the latest developments in carbon research described by Kroto, novel results are continuing to be reported on the more 'conventional' post-fullerene carbon forms, such as the nanotubes first discovered by S. Iijima last year (*Nature* 354, 56–58; 1992). Within months, T. W. Ebbesen and P. M. Ajayan had discovered how to make these nanotubes in large quantities (*Nature* 358, 220–222; 1992). And L. A. Chernozatonskii and colleagues have succeeded in making ordered arrays of these tubules, as shown in the micrograph (Z. Ya. Kosakovskaya *et al.* *JETP Lett.* 56, 26–30; 1992). The authors synthesized the material by electron-beam evaporation of graphite in a high vacuum (10^{-8}



atmospheres). The electron microscopy reveals that the texture formed on the surface comprises carbon filaments (themselves made of several tubules) 5 nm across and gathered into cables 10–100 nm long. By tilting the substrate, the authors found they could change the orientation of the carbon texture, leading them to propose that the pressure in the carbon plasma reflected from the substrate surface builds up enough for fullerene growth to be nucleated and for the tubes to grow to great length before being 'capped off' at the top. The individual tubes were about 0.8 to 1.1 nm in diameter, corresponding to the narrowest and second narrowest possible structures, the authors say. Conductivity measurements suggest that the film is semiconducting or semimetallic (individual tubes' conductivity can depend enormously on their diameter, according to calculations). The authors were impressed by the texture's hardness which exceeds that of the toughest alloys they used as substrates. (Scale bar is 100 nm.) R.P.

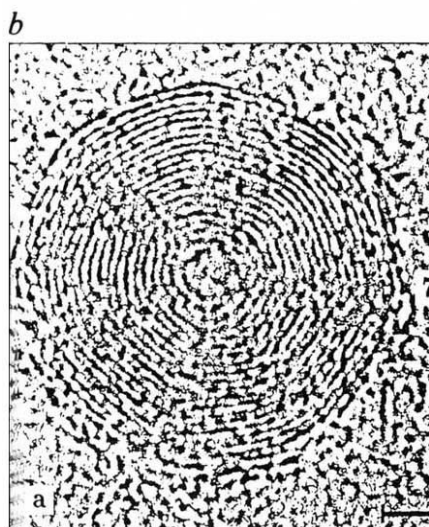
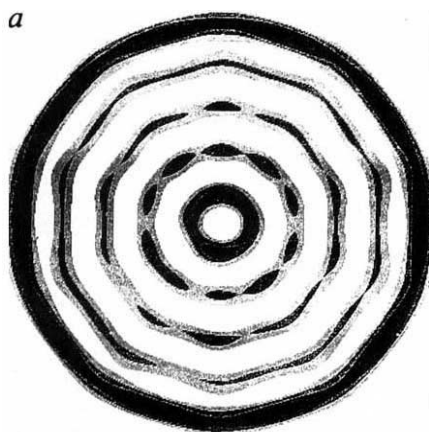


FIG. 2 *a*, Simulated transmission electron micrograph of giant fullerene concentric shells^{10,11}, and *b*, real micrograph of nested graphite shells, by Iijima. Compare with Ugarte's micrograph on this issue's cover.

which carbon atoms can rearrange themselves. More careful study is required to understand whether a nucleus is involved, but there is evidently an ordering process in which successive outer shells assemble around inner ones. Ugarte's very large structures are more spherical than Iijima's polyhedral particles, perhaps because they are not perfectly closed shells. The carbon nanotubes made last year by Iijima⁹ can be considered as elongated analogues of these spheroidal structures. But it is interesting that Ugarte's study indicates

that elongated (amorphous?) structures appear to evolve or anneal into spheroidal ones. This suggests that spinning of carbon nanotubes by high-temperature fusion of graphite might be difficult to achieve directly.

Ugarte's experiments may also take us back to the original motivation behind the carbon-cluster experiments that revealed fullerene — the question of what form carbon takes in interstellar space. The results add further evidence to conjectures, based on Iijima's original re-

sults, that some carbon particles in space might consist of concentric-shell graphite structures, perhaps surrounded by amorphous outside layers.

Nevertheless, the most interesting question is whether, 500 years after Columbus reached the West Indies, flat carbon has gone the way of the flat Earth. □

H. W. Kroto is in the School of Molecular Sciences, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

G PROTEINS

Visual differences

Patrick Casey

NEWS from the G-protein front continues to stream in, the latest dispatches coming from Kokame *et al.* on page 749 of this issue¹ and from Neubert *et al.* in the *Journal of Biological Chemistry*². They report that an unexpected phenomenon — heterogeneous acylation of the α -subunit amino terminus of the retinal G protein, transducin — may exert control over the protein's activity. Here we have the opening of what should be a highly interesting story.

Many proteins involved in membrane signalling are modified by the covalent attachment of lipids, and the effects of such modifications on the pathways controlled by the superfamily of GTP-binding regulatory proteins (G proteins) have come under particularly close scrutiny. For instance, attachment of isoprenoid lipids is required for many of these molecules to function^{3,4}. Another well-studied example is protein *N*-myristoylation (attachment of a 14-carbon saturated fatty acyl group)⁵, which often modifies the α -subunits of the heterotrimeric G proteins. These heterotrimeric G proteins play a central role in signalling across cell membranes, as discussed by Bourne and Stryer in a recent News and Views article⁶. The prevailing dogma had it that this form of protein acylation was specific for myristate, because no other fatty acyl groups had been found to be linked through an amide bond to the amino-terminal glycine residue of proteins. But with the appearance of the papers by Kokame *et al.* and Neubert *et al.* this dogma now lies in ruins.

The new studies arose out of attempts to explain a paradox concerning G-protein myristoylation. The photoreceptor G protein involved in visual signal transduction, known as transducin or G_t , was unique among the group of pertussis toxin-sensitive G proteins in that no myristoyl group was found on its α -subunit ($G_{t\alpha}$) by chemical analysis. In

contrast, the related $G_{i\alpha}$ and $G_{o\alpha}$ subunits purified from brain were myristoylated⁷. Metabolic labelling studies conducted in cells expressing $G_{t\alpha}$, though, revealed that the protein could incorporate [³H]myristate⁸, indicating that its amino terminus contained the correct signal to direct myristoylation. Is the environment in retinal cells not conducive to myristoylation of $G_{t\alpha}$, or is there some other modification on the amino-terminus of $G_{t\alpha}$ that prevents its myristoylation? Elements of both these possibilities turn out to be involved in the resolution of the paradox.

Kokame *et al.*¹ and Neubert *et al.*² found that $G_{t\alpha}$ purified from retinal rods is not a single entity but a small family of proteins. Mass spectrometry revealed that there are four members of this family, each with the same amino-acid sequence but containing one of four distinct fatty acyl groups on the amino terminus (see figure). One member contains the myristoyl group (termed C14:0, which denotes a chain length of 14 carbons and no double bonds), though this myristoylated $G_{t\alpha}$ is the minor species. The amino termini of the other three $G_{t\alpha}$ species contain (in order of abundance) a lauryl (C12:0) group or one of two unsaturated acyl groups (C14:2 or C14:1). These three acyl groups are all less hydrophobic than C14:0, and the unsaturated ones would have conformational restrictions. Consequently, the $G_{t\alpha}$ molecules they modify could have significantly altered properties compared to their myristoylated equivalents. Kokame *et al.* provide some intriguing data which suggest that this is the case.

The amino-terminal myristoylation of specific G-protein α -subunits not only influences the membrane association of these polypeptides⁸, but also increases the affinity of the α -subunit for its cognate $\beta\gamma$ -subunit complex in the absence of membranes⁹. Subunit interactions are

1. Ugarte, D. *Nature* **359**, 707–709 (1992).
2. Krättschmer, W., Lamb, L. D., Fostiropoulos, K., Huffman, D. R. *Nature* **347**, 354–358 (1990).
3. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
4. Jones, D. E. H. *New Sci.* 245 (3 November 1966).
5. Iijima, S. *J. Cryst. Growth* **5**, 675–683 (1980).
6. Kroto, H. W. & McKay, K. G. *Nature* **331**, 328–331 (1988).
7. Zhang, Q. L. *et al.* *J. Phys. Chem.* 525–528 (1986).
8. Maryuma, S., Anderson, L. R. & Smalley, R. E. *Rev. sci. Instrum.* **61**, 3686–3693 (1990).
9. Iijima, S. *Nature* **354**, 56–58 (1991).
10. Kroto, H. W. *JCS Farad. Trans.* **86**, 328–331 (1988).
11. McKay, K. G., Wales, S. W. & Kroto, H. W. *JCS Farad. Trans.* (in the press).

RÉSUMÉ

Memory logs

H. KUMAGAI and Y. Fukao have turned petrified trees to good use (*Geophys. Res. Lett.* **19**, 1859–1862; 1992) in reconstructing events at a now extinct volcano in central Honshu. Twenty million years ago the slopes of the volcano were violently deforested in a sequence of hot avalanches and explosions. The tree trunks, caught up in the deposits and then fossilized by silica mineralization, are so well preserved that the volcanologists were able to count and correlate tree-ring sequences from many logs and deduce a year-by-year growth pattern for the population. They could recognize at least five eruptions when some trees died while others survived, in a sequence lasting 17 years — an unprecedented time resolution for ancient volcanic activity.

Dominant domain

ANKYRIN is a protein that binds to a variety of other proteins — the red cell membrane anion channel and spectrin, thus attaching the membrane skeletal network to the membrane, tubulin, vimentin and the Na^+/K^+ -ATPase of epithelial cells. Ankyrin contains three recognizable domains, one of which regulates the others. In red cells the protein occurs as a set of isoforms, some of which at least are splicing variants; in one the regulatory domain is deleted, and binding to other proteins is thereby enhanced. L. Davis *et al.* (*J. biol. Chem.* **267**, 18966–18972; 1992) have now shown that the isolated regulatory domain not only attaches to the truncated variant of ankyrin, but also non-competitively moderates its binding to the anion transporter. Why such fine-tuning is needed for a protein that had been thought of as no more than a kind of clothes-peg is now open to conjecture.

New for old

TEMPERATURE fluctuations in the microwave background, whose detection by the COBE satellite was so widely reported earlier this year, imply such small primordial density fluctuations that it is already hard for theoretical cosmologists to manufacture galaxies and clusters in the time available. C. Hogan is now trying to make the job even harder (*Astrophys. J.* **398**, L77–L80; 1992). According to Hogan, it is possible that much of the COBE signal is actually of rather recent vintage. Hot ionized gas associated with galaxy superclusters is not quite transparent to the passage of microwave photons, so can induce extra temperature irregularities. But then the primordial fluctuations must be smaller still, which makes galaxy formation harder yet. . . which leaves us wondering where the superclusters and their hot gas could have come from.

a crucial component of G-protein action, because subunit dissociation upon activation of the G protein by GTP binding is a requirement for efficient signal transmission⁶. A unique property of G_{ta} is that, unlike other α -subunits, it does not require detergent for its release from membranes. So a major role of acylation of G_{ta} may be to control subunit interactions rather than membrane association.

Kokame *et al.* synthesized acylated peptides corresponding to the amino terminus of G_{ta} and, in *in vitro* measure-

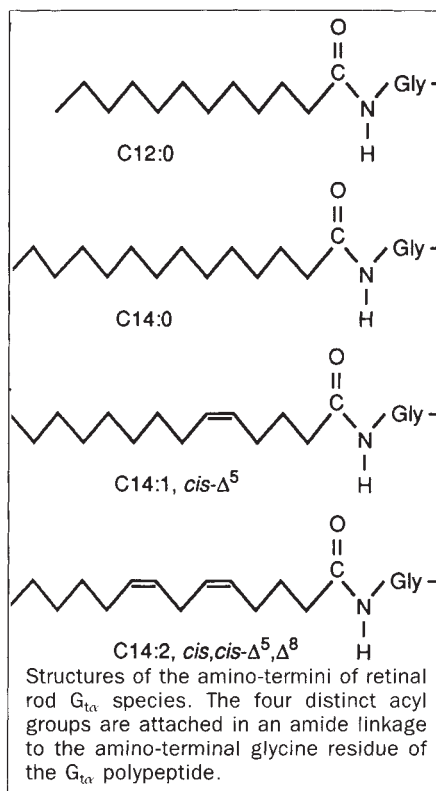
supports the view that heterogeneous acylation is retinal-specific.

One possibility is that this heterogeneous protein acylation in the retina reflects the relative abundance of acyl-CoA substrates available to the myristoyl-CoA:protein *N*-myristoyltransferase (NMT), which is the enzyme thought to be responsible for this modification. Another is that there is a retinal-specific NMT with unique selectivity for acyl CoAs. The first possibility seems most likely, because the characterization of NMT has revealed that it can accept a variety of acyl-CoA species as substrates, including those with C12:0 and unsaturated acyl groups¹¹, even though the myristoyl group is the one that (until now) was always found on modified proteins.

So it seems likely that the heterogeneous acylation of G_{ta} will reflect retinal-specific production of acyl-CoA species. Analysis of other retinal counterparts of known myristoylated proteins, particularly those not involved in signalling in the rod outer segment, should settle the matter, as they should also contain the same mixture of acyl groups if retinal acyl-CoA levels are responsible. Another approach is to determine if expression of G_{ta} or recoverin in a heterologous cell type produces a protein with only C14:0 on its amino terminus — in the previous study with G_{ta} only the fate of the labelled myristate was followed, and other acyl groups attached to the protein might not have been detected.

A host of issues now present themselves. How, for example, are acyl-CoA levels in cells determined and what is the effect of those levels on protein acylation? And what is the significance of producing several forms of an acylated protein from a single translation product? □

Patrick Casey is in the Section of Cell Growth, Regulation and Oncogenesis and the Department of Biochemistry, Duke University Medical Center, Box 3686, Durham, North Carolina 27710, USA.



ments of α - $\beta\gamma$ interactions, found that the C12:0-modified peptide was a much less potent competitor of this interaction than a myristoylated peptide. This observation suggests that the G_{ta} modified by C12:0 interacts more weakly with $\beta\gamma$ than one which is modified by C14:0. This weaker interaction could lead to a more rapid dissociation of the G_i oligomer upon light stimulation than would be possible if the α -subunit was modified by C14:0, and the authors speculate that this property could contribute to the rapidity of the visual signalling response compared with that of other systems controlled by G proteins.

If modification of G_{ta} in retinal rods by acyl groups less hydrophobic than myristate determines the rapidity of visual signalling, the question of how this modification arises becomes an important one. Notably, recoverin, another retinal protein involved in visual signalling, has also recently been reported to contain the same acyl groups at its amino terminus as G_{ta} (ref. 10). This finding

- Kokame, K., Fukada, Y., Yoshizawa, T., Takao, T. & Shimonishi, Y. *Nature* **359**, 749–752 (1992).
- Neubert, T. A., Johnson, R. S., Hurley, J. B. & Walsh, K. A. *J. biol. Chem.* **267**, 18274–18277 (1992).
- Clarke, S. A. *Rev. Biochem.* **61**, 355–386 (1992).
- Inglese, J., Koch, W. J., Caron, M. & Lefkowitz, R. J. *Nature* **359**, 147–150 (1992).
- Gordon, J. I. *et al. J. biol. Chem.* **266**, 8647–8650 (1991).
- Bourne, H. R. & Stryer, L. *Nature* **358**, 541–543 (1992).
- Buss, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 7493–7497 (1987).
- Mumby, S. M., Huekeroth, R. O., Gordon, J. I. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 728–732 (1990).
- Linder, M. E. *et al. J. biol. Chem.* **266**, 4654–4659 (1991).
- Dizhoor, A. M. *et al. J. biol. Chem.* **267**, 16033–16036 (1992).
- Kishore, N. S. *et al. J. biol. Chem.* **266**, 8835–8855 (1991).

Long time series reveal density dependence

H. C. J. Godfray and M. P. Hassell

A PAPER in the latest issue of *Journal of Animal Ecology*¹ makes an outstanding contribution to one of the oldest debates in ecology. Woiwod and Hanski report on the analysis of a unique data set of long-term fluctuations in nearly 6,000 populations of aphids and moths. They find that density dependence is pervasive and common among these populations, thereby confounding those who have claimed density dependence to be rare, absent or impossible to detect in populations of insects.

It is a logical necessity that any population of plants or animals that persists in the environment must experience some form of density-dependent feedback on net population growth as population densities rise (either a reduction in fecundity or immigration, or an increase in mortality or emigration). But although the notion has lain at the heart of population biology for most of this century, density dependence has proved notoriously difficult to detect.

Controversy over the action of density dependence has been particularly intense among insect ecologists. For other taxa, density dependence is so obvious that it excites little comment. Many plants and sessile animals, for example, clearly compete for space at high population densities, as do territorial birds; likewise, herbivorous mammals often exhaust their food supply when numbers are high and suffer unambiguous density-dependent mortality. In contrast, insect populations typically seem to fluctuate widely and wildly, and there is little obvious relationship between population density and the strength of mortality. The failure to detect frequent density dependence in several surveys of available data² has led to the suggestion that the densities of most insect populations wander in a random walk for many generations, only occasionally reaching numbers that are sufficiently high for density dependence to operate³.

Critics of this view make two points^{4,5}. First, and most importantly, there are formidable statistical problems in the detection of density dependence in short runs of population densities. Many of the techniques used in the past are now accepted to be flawed and the methods currently favoured, those of Bulmer⁶ and Pollard *et al.*⁷, require longer runs of data. Quite possibly the rarity of density dependence may be just a reflection of the paucity of suitable data. Second, the data that are available are heavily biased towards species of economic importance,

and thus contain a disproportionate number of outbreak pests, which by definition have unusual forms of population regulation. To resolve this problem, density dependence needs to be studied in long runs of data of species chosen without a particular bias.

Rothamsted Experimental Station in the south of England has for the past 30 years conducted a survey of the abundances of aphids and moths at sites throughout Great Britain. Moths are sampled using ultraviolet light traps, and aphids by suction traps which filter the air for small flying insects. Woiwod and Hanski studied 5,715 time series of 94 species of aphids and 263 species of moths; all time series were more than ten years in length, and many more than 20 years. Using the Bulmer test, they found significant density dependence in

81% of the aphid and 47% of the moth time series; using the test of Pollard *et al.*, the equivalent figures were 69% and 29% respectively. If analyses were restricted to time series of 20 or more years, the frequency of density dependence increased further: the more conservative Pollard test found density dependence in 84% of the aphid time series and 57% of the moth. Some tests perform less well when populations show long-term changes in characteristic abundance, for example in response to habitat change, and when Woiwod and Hanski excluded time series with a significant linear trend in density over time the frequency of density rose yet higher. Again using the more conservative Pollard test, 87% of aphid time series, and 67% of moth time series, showed evidence of density dependence.

The results mean that, when suitable data are available, insect ecologists should be able to detect density dependence from population time series. Although some populations will experience only rare episodes of density dependence and otherwise fluctuate for

A long shot at the Kuiper belt

HAVING discovered the most distant known object in the Solar System, also possibly the first known member of the much hypothesized Kuiper belt of comets, David Jewitt and Jane Luu have earned the right to a little speculation. Jewitt duly offered some thoughts about the Kuiper belt at the meeting in Munich last week of the American Astronomical Society's Division of Planetary Sciences. After two sightings of 1992QB1, now known as Smiley after le Carré's character, 3½ weeks apart, Jewitt and Luu can say the

expected from Kepler's law for bodies so far away. Jewitt suggested that the velocity dispersion in the Kuiper belt, indicated by this small difference, is very low, not much more than the escape velocity of the member bodies. This is in contrast to the asteroid belt, where the velocity dispersion is very large; whereas the bodies there are the products of disruptive collisions, collisions between members of the Kuiper belt would apparently lead to aggregation. One participant pointed out that a low velocity

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

1992QB1 (indicated) was discovered during a painstaking search of the sky for faint objects in the Kuiper belt. Candidate comets need to be distinguished by their slow motion from the plethora of galaxies that litter the sky. The trail to the right is due to an asteroid, which moves relatively quickly.

object is 42 astronomical units away, just beyond Pluto's mean orbit. Its detection after a careful survey of 0.7 square degrees of the sky (since 1987) implies that there are 10,000 similar bodies (around 240 km in diameter) in the Kuiper belt, assuming the belt is 20° deep, having in total the same mass as Pluto. Estimates by M. Duncan, T. Quinn and S. Tremaine, based on the number of short-period comets which are thought to originate in the Kuiper belt, yielded a similar conclusion. Also, the velocity of 1992QB1 is not much different from that

dispersion would conflict with the large assumed depth of the Kuiper belt, so leaving the story a bit incomplete. On the positive side, 1992QB1 is very red, like the recently discovered distant asteroidal object Pholus, which seems to have been ejected from the Kuiper belt not long ago. Although tracking until January will be needed to confirm that the orbit is as expected for a Kuiper-belt member and although speculation on a sample of one is a dangerous thing, it is a lot better, as Jewitt pointed out, than speculation on a sample of nought.

R.P.

Jane Luu

many generations as a random walk process, the analysis reduces the likelihood that the random walk is a good model of the dynamics of many insect populations. Woiwod and Hanski's work also once again emphasizes the crucial importance of long-term data collection in ecology.

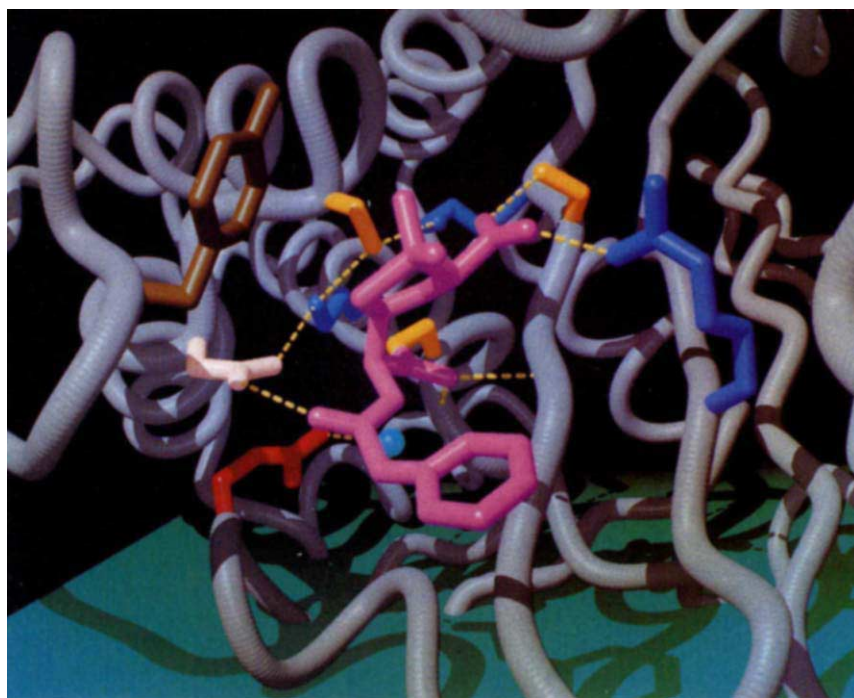
In the controversy over insect population regulation, it is the more theoretically inclined ecologists who have tended to argue that density dependence is common, whereas many empiricists have been particularly attracted by theories emphasizing the unpredictability of natural systems. Empiricists look at the deterministic outcome of many theoretical models incorporating simple forms of density dependence and, knowing the world is not like that, take comfort in the unpredictable outcomes of populations modelled as random walks, and in suggestions that density dependence operates only very infrequently. Theoreticians, on the other hand, confident that density dependence must be present in persistent populations, sniff apostasy in attempts to play down the frequency with which it operates in natural populations.

The division between the two camps is probably less than it seems. There is a continuum of types of density dependence from forms that increase in severity in a simple manner across population densities, to forms that operate only over a sharp threshold⁸. Similarly, the relationships between density dependence and population density may be relatively constant, or the relationships may contain much noise and scatter. Woiwod and Hanski's results are encouraging in that they suggest that the type of density dependence occurring in natural populations of insects may not be as intractable to study as many have feared. The detection of density dependence is however only one step, albeit an important step, in understanding the dynamics of insect populations. We still need to know the biological basis of density dependence, how its strength varies with population density, and how it interacts with other sorts of mortality to determine observed population fluctuations. □

H. C. J. Godfray and M. P. Hassell are in the Department of Biology and the NERC Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

1. Woiwod, I. P. & Hanski, I. J. *anim. Ecol.* **61**, 619–630 (1992).
2. Stiling, P. J. *anim. Ecol.* **57**, 581–593 (1988).
3. Den Boer, P. J. *Oecologia* **86**, 484–491 (1991).
4. Hassell, M. P., Latto, J. & May, R. M. J. *anim. Ecol.* **58**, 883–892 (1989).
5. Holyoak, M. & Lawton, J. H. *Oecologia* (in the press).
6. Bulmer, M. G. *Biometrics* **31**, 901–911 (1975).
7. Pollard, E., Lakhani, K. L. & Rotheray, P. *Ecology* **68**, 2046–2055 (1987).
8. Strong, D. R. *Trends Ecol. Evol.* **1**, 39–42 (1986).

Drug ring bust



THIS futuristically colourful picture tells quite a story. It shows the mechanism by which the enzyme β -lactamase nullifies the antibiotic effects of penicillin, as reported by Strynadka *et al.* on page 700 of this issue.

Penicillins (and the cephalosporins) are β -lactam antibiotics which interfere with bacterial cell-wall synthesis, and consequently kill actively growing and dividing bacteria. The peptidoglycan network of the bacterial cell wall is formed by the synthesis of linear polysaccharide chains and the cross-linking of these chains with short peptides. It is this latter process of transpeptidation that is inhibited by the β -lactam antibiotics; they react irreversibly with the active site of the transpeptidases that catalyses the cross-linking.

But not all pathogenic bacteria are sensitive to these antibiotics; penicillin-resistant strains often harbour plasmids (so-called bacterial drug resistance factors) that specify β -lactamase enzymes which break down the β -lactam ring of the penicillin. The hydrolysis of the amide bond in the ring yields penicilloic acid which is devoid of antibiotic activity.

Strynadka *et al.* now provide clear insight into how these antibiotics are recognized and destroyed. They present the high-resolution crystal structures of the *Escherichia coli* RTEM class A β -lactamase and a mutant version of the same enzyme defective in deacylation, both on its own and complexed with its natural substrate, penicillin G.

Like the serine proteinases and the cell-wall synthesizing enzymes they protect from antibiotics, the β -lactamases have at the centre of their catalytic machinery a serine residue that functions as a nucleophile, attacking the carbonyl

carbon of the β -lactam ring. Unlike the serine proteinases, however, distinct residues are implicated in the acylation and deacylation steps, and the two steps occur by almost completely different mechanisms.

The amino-acid residues important in catalysis are shown in this depiction of the acyl-enzyme (dashed lines are hydrogen bonds). The penicilloyl intermediate of penicillin G (pink) is covalently attached to the enzyme (α -carbon backbone in grey) via the nucleophile serine 70 (yellow, bottom).

Lysine 73 (blue tube to the left of the substrate) functions as a general base in assisting in the initial nucleophilic attack, and in combination with serine 130 (yellow, top left), in protonating the leaving group nitrogen of the β -lactam ring to form the acyl-enzyme intermediate. Glutamine 166 (red) is the general base in the deacylation reaction, assisting the nucleophilic attack by a water molecule (aquamarine sphere) on the ester carbonyl carbon of the acyl-enzyme intermediate.

The remaining residues — tyrosine 105 (brown), arginine 244 (blue, right), asparagine 132 (peach) and serine 235 (yellow, top right) among them — are involved in substrate binding.

β -lactamase enzymes are produced by many bacteria. Increased understanding of the way they work has allowed the development of resistant antibiotic analogues (for example ampicillin) and has led to the discovery of naturally occurring lactamase inhibitors (for example clavulanic acid). Strynadka *et al.* have added an extra layer of detail to knowledge of the catalytic mechanism concerned, which will no doubt guide attempts to design better inhibitors. G.R.

A brief history of pulsar time

R. D. Blandford

RADIO pulsars were discovered¹ because their sources, spinning neutron stars, keep time with extraordinary precision. Indeed, some rival the very best atomic clocks. But they are not perfect clocks. Allowance must be made for their steady deceleration due to the action of an electromagnetic torque; and some undergo sudden, unpredictable increases in their frequencies, known as glitches. One of the best studied pulsars is PSR0531+21, more famous as the central pulsar of the Crab Nebula — the relic of a supernova explosion observed by Chinese astronomers in AD 1054. A large glitch on 29 August 1989 shortened this pulsar's 33-millisecond period by about 3 nanoseconds, a much bigger jump than in previous glitches. On page 706 of this issue², Lyne, Graham Smith and Pritchard present an analysis of the times of arrival of pulses from this pulsar which suggests a physical description of the interior of the neutron star that is both more complicated and more intriguing than any that has been seen so far.

By now, many pulsars have been found to undergo glitches. The most capricious example is the Vela pulsar, PSR0833-45, whose frequency has jumped by a few parts per million on eight occasions over 20 years. After each glitch, there was a temporary increase in the rate of slowing down so that the star appears to be trying to recover the frequency it would have had, had it not glitched. The best scrutinized Vela glitch, which occurred on Christmas Eve in 1988, had at least three distinguishable recovery timescales, the shortest being 10 hours and the longest being 3 months³.

However, no two glitching pulsars behave similarly. For example, PSR0355+54, which underwent the largest glitch of all in 1986, exhibited almost no recovery⁴. In addition, PSR1737-30, which was first found to glitch in 1987, was most prolific, performing four encores over the next two years⁵. Now we can add the Crab pulsar to this list. Its recent glitch is unique in revealing three separate recoveries, of which two are negative so that the frequency slowly increases relative to pre-glitch expectations. Furthermore, the steady deceleration of the Crab pulsar is apparently slightly faster now, in contrast with other glitching pulsars, which generally return to their pre-glitch slowing-down rate.

Faced with this variety, it should come as no surprise to learn that theoretical astrophysicists have had a very hard time

persuading observers (and each other) that they understand glitches. Most published explanations have failed to predict the behaviour of future glitches. (One anonymous physicist appeared so confident that the first Vela glitch was a singular event that he was offered a Faustian bargain to live until the next. Fortunately, he had the wisdom to decline the offer.) As is so often the case in theoretical astrophysics, two approaches have been followed, working inductively, backwards from the data and deductively, forwards from physical theories of neutron-star matter.

The simplest, phenomenological interpretation of the results of Lyne *et al.* is that the neutron star contains four distinct zones just as the interior of the Earth is known to comprise a solid core, a liquid outer core, a solid mantle and a crust. In the neutron star, the four zones can spin with different frequencies, but just before the glitch, they are in synchrony. During the glitch, the star undergoes some structural change and, according to the principle of conservation of angular momentum, all four zones start spinning with different frequencies.

The pulses that we observe are generated in the magnetosphere outside the star and, in the case of the Crab pulsar, we must say that this is attached to the third-fastest component. What happens next, in this interpretation, is that frictional forces act between the different zones, rather like the clutch in a gearbox. The zone that is linked to the magnetosphere, and is responsible for the pulsar clock, is first spun up by the friction from a faster zone until both components have a common rotational frequency. Then these two components are decelerated by a slower component and, finally, the remaining, faster zone exerts a diminishing spin-up torque that leaves the components in synchrony once more. This model describes the timing record.

The alert reader might ask at this point: how can one zone be in contact with three others? After all, the liquid core of the Earth naturally has two boundaries. This brings us to the deductive model. The answer is believed to involve a remarkable property of the neutrons in the interior of the star — they are thought to be superfluid⁶. This means that in certain density ranges, the neutrons can flow through the crystalline lattice formed by the ions with relatively little resistance, just like very cold liquid helium in the laboratory.

There is a problem, though — perfect

superfluids are not supposed to rotate. The way out is that the rotation (strictly a quantity known as the vorticity) is concentrated in a bundle of parallel vortex lines (separated from each other by about 20 micrometres in the case of the Crab pulsar), whose filamentary cores contain normal, non-superfluid neutrons⁷. As the star slows down fewer of these vortex lines are needed and so they ought to move away from each other. However, there is a second complication. The vortex lines may get pinned by the ions in the lattice. This means that as the magnetosphere and the ions, which are linked by magnetic field, slow down, the neutron superfluid may keep on spinning with the same frequency.

When the frequency difference between the star and the lattice gets too large, the stresses in the crust become so great that something has to give and there is an exotic form of starquake. Some older (and perhaps colder) neutron stars, like PSR0355+54, may behave like this. Alternatively, after a glitch in a young (and perhaps hotter) pulsar like the Crab or Vela, the vortex lines may creep slowly through the lattice, reducing the velocity difference between the neutrons and the ions, and exerting the frictional drag that characterizes the recovery⁸. If this is what is happening, then in order to account for the behaviour of the Crab pulsar, some superfluid zones must be left by the glitch spinning faster than the magnetosphere.

Undoubtedly, more detailed models will accommodate the slow spin-ups in the Crab pulsar. However, as Lyne *et al.* remark, there is a fairly fundamental difficulty if the pulsar has undergone glitches like this one ever since its birth, as the moment of inertia of the mobile neutron superfluid would have to be as large as 5 per cent of that of the whole star, far larger than is allowed by models of neutron star interiors. Despite the observational and theoretical progress that has already been made, understanding and predicting neutron starquakes is proving to be no easier than dealing with their terrestrial counterparts. □

R. D. Blandford is in the Department of Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125, USA.

- Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F. & Collins, R. A. *Nature* **217**, 709–713 (1968).
- Lyne, A. G., Graham Smith, F. & Pritchard, R. S. *Nature* **359**, 706–707 (1992).
- Flanagan, C. S. *Nature* **345**, 416–417 (1990).
- Lyne, A. G. *Nature* **326**, 569–571 (1987).
- McKenna, J. & Lyne, A. G. *Nature* **343**, 349–350 (1990).
- Baym, G., Pethick, C. J., Pines, D. & Ruderman, M. *Nature* **224**, 872–874 (1969).
- Feynman, R. P. *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, 1972).
- Alpar, M. A., Pines, D. & Cheng, K. S. *Nature* **348**, 707–708 (1990).

Second gorilla or third chimp?

David Dean and Eric Delson

WHAT is the origin of and the relationships within the primate lineage including gorillas, chimpanzees and humans? That nexus of questions is among the most compelling in anthropology, and David Begun¹ has brought two new hypotheses to bear on it.

Begun's analysis of recently recovered fossils of *Dryopithecus*, a European ape dating to 13–9 million years ago, leads him to suggest, first, that this group of species shares with the gorilla a complex of facial and dental features which are ancestral for the African-ape and human subfamily, Homininae. Even more intriguing is Begun's second suggestion, that features shared by chimpanzees and the extinct 'early human' *Australopithecus* link the chimp and the human sublineages within Homininae to the exclusion of the gorilla. This second hypothesis has long been championed by some molecular systematists^{2,3}, although others have disagreed⁴. If Begun's work is correct, it would constitute a major contribution to palaeoanthropology (see also ref. 5). But, stimulated by his proposals, we have examined them in light of other recently described fossils and come to somewhat different conclusions.

Although several species of *Dryopithecus* are now represented by partial craniofacial remains, the 8–10 million-year-old Greek ape *Graecopithecus* (also called *Ouranopithecus*^{6,7}) is known by most of a male face (Fig. 1a) and numerous jaws. These fossils show greater similarity to *Pan*, *Australopithecus* and especially *Gorilla* (Fig. 1b) than does *Dryopithecus*. In our opinion, *Graecopithecus* is a far stronger candidate for an early member of Homininae, if not of the gorilla subclade.

Shared features

Begun lists a variety of character states which link *Dryopithecus* to hominines, and he specifically indicates that *Graecopithecus* shares all that could be seen on the known fossils, but he refrains from further interpretation of the Greek face. He relates several of these shared features to klinorhynch (the downward bending of the face on the brain case⁸); we agree, and explicitly suggest that an increase in klinorhynch was related to the origin of Homininae. Three further characteristics of the upper dentition are shared by all hominines, including both European fossils: the incisor crowns are oriented vertically to take advantage of what was probably a horizontally oriented posterior temporalis muscle like that seen in the African apes and humans; the canines, especially their shear

facets, are in line with postcanine teeth rather than laterally rotated; and the lingual cusps on the molars tend to be posterior to their buccal neighbours — species with less klinorhynch or airorhynch (upwardly bent) faces tend to have transversely aligned cusps and squared-off molars.

Begun also does not take into account the point that *Graecopithecus* presents more extreme klinorhynch than does *Dryopithecus*. For example, the brow ridge of the new Greek face⁷ is continuous as in *Dryopithecus*, but it projects at glabella (in the midline) to a degree otherwise found only in later hominines. In addition, *Graecopithecus* resembles *Gorilla* more than other

hominines in its superolateral brow and lateral orbit margin shape; the sharply angular muzzle shape with supra-alveolar waisting; anterior nasal bone projection; and (less definitively due to damage) in the development of the frontal trigone, the area just behind the brow ridge bounded by the anterior temporal line.

Different interpretations

So why does Begun not include *Dryopithecus*, much less *Graecopithecus*, squarely in the Homininae (see Fig. 2a)? The reason is that he interprets a somewhat elongated premaxilla and larger maxillary sinus as possibly having characterized the hypothetical common ancestor of living great apes, excluding *Dryopithecus*. In our opinion, both features are probably size related, and Begun himself realizes that the naso-alveolar elongation in the airorhynch

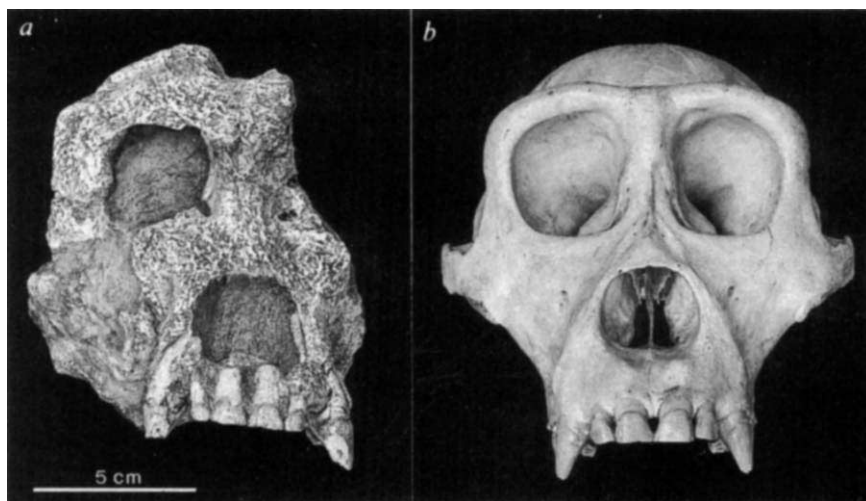


FIG. 1 Frontal view (alveolar plane horizontal) of, a, *Graecopithecus freybergi* (cast of male face from Xirochori?) and, b, female *Gorilla gorilla*. Similarities between the two species are discussed in the text. Photo Lorraine Meeker/Eric Delson.

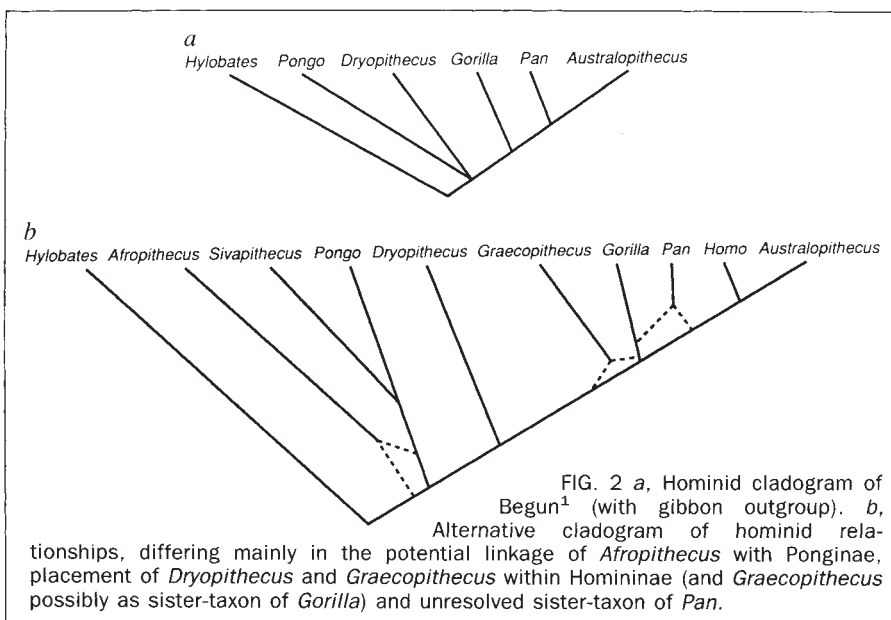


FIG. 2 a, Hominid cladogram of Begun¹ (with gibbon outgroup). b, Alternative cladogram of hominid relationships, differing mainly in the potential linkage of *Afropithecus* with Ponginae, placement of *Dryopithecus* and *Graecopithecus* within Homininae (and *Graecopithecus* possibly as sister-taxon of *Gorilla*) and unresolved sister-taxon of *Pan*.

orangutan clade (Ponginae) is convergent with that seen in the klinorhynch *Pan*.

Our cladogram (Fig. 2b) presents a rather different interpretation of hominid relationships. The Asian fossil ape *Sivapithecus* is airohynch and clearly a pongine craniodentally, although it lacks several postcranial features that have long been interpreted to characterize all hominids⁹. Moreover, the 17-million-year-old face of Kenyan *Afropithecus*¹⁰ is also airohynch, as demonstrated for example by its upturned orbits and premaxilla, procumbent incisors and molar cusps aligned transversely. If *Graecopithecus* is on the gorilla subclade, that lineage dates to at least 9 million years ago.

Begun¹ also proposes morphological support for a link between chimpanzees and the human subclade, which he claims share derived features of the frontal bone, subnasal region, upper molars and molar proportions not seen in *Gorilla*. This aspect of the work is less detailed and less convincing; if some of these states are ancestral or convergent, the case would be greatly weakened. For example, the subnasal elongation and concomitant increase in maxillary sinus size in *Pan* and *Australopithecus* may be convergent, as it is not associated in the latter with the extreme klinorhynch shared by *Pan* and *Gorilla*.

Moreover, Begun reasonably questions the evidence from tooth enamel thickness linking *Pan* and *Gorilla*, but also rejects the uniqueness of their shared knuckle-walking morphology. He argues that the latter is a single functional complex and thus effectively of lesser weight than the multiple dento-facial characters which oppose it. But the question of how to weight a complex of perhaps dozens of individual 'characters' is far from resolved: if we weight each component at half a 'full' character, the chimp-gorilla link 'wins'. We consider that the trichotomy remains unresolved.

Diamond¹¹ has referred to humans as "the third chimpanzee" in recognition of our close biomolecular similarity to the

two living species of *Pan*. Begun's article supports this view, but we are not entirely convinced that the closest genealogical relative of chimpanzees is the human clade rather than the morphologically similar gorilla. We await further specimens of the *Graecopithecus* cranium to assess variation, as well as deeper analysis to confirm that its similarities to *Gorilla* are derived within Homininae. At the least, however, the reinterpretation

here of *Graecopithecus* suggests that it is a fourth lineage within the Homininae, if not a 'second gorilla'. □

David Dean is in the Department of Cell Biology, New York University Medical Center, New York, New York 10016, USA. Eric Delson is in the Department of Anthropology, Lehman College and the Graduate School, City University of New York, Bronx, New York 10468, USA.

EARTHQUAKES

Weakness amidst strength

Christopher H. Scholz

WITH growing evidence that the San Andreas fault is weak relative both to the surrounding rock and to the frictional strength of appropriate geological materials, a minor cottage industry has sprung up in the past few years in trying to explain this apparently paradoxical behaviour. There are two general types of model: those that propose that faults are lined with some material, such as clay, with an intrinsically low friction coefficient; and those that invoke high pore-fluid pressure to reduce the effective normal stress, which generates the frictional resistance. Geological studies have cast the former into doubt because low-friction materials are not found to be ubiquitous within faults¹, and most recent efforts have concentrated on models in which high pore pressures are localized within the fault, owing to the contrast of hydraulic and mechanical properties of the fault zone materials with that of the surrounding rock²⁻⁴. On page 687 of this issue⁵, Sleep and Blanpied have taken this latter class of models much further, in proposing a means in which pore pressure within the fault fluctuates during the earthquake cycle as a natural consequence of the cycle itself.

Friction experiments

Their model was inspired by the friction experiments recently performed by Blanpied, along with Lockner and Byerlee⁶, which revealed anomalously low frictional sliding resistance at high temperature and pressure in a gouge (rock-powder) filled sawcut that contained an initial pore pressure. Time-dependent compaction of the gouge raised the entrained pore pressure, which then became sealed in by dissolution at the contact with the surrounding rock. The pore pressure within the gouge was increased by about a factor of three by this mechanism, reducing the sliding resistance without altering the intrinsic coefficient of friction.

Sleep and Blanpied now codify this

mechanism using a model of a fault containing a core of porous material which flows ductilely, both in shear and compaction modes. As the fault core, which they assume is sealed from the surrounding rock, is slowly compacted during the interseismic period, the pore pressure progressively builds up in the fault core, eventually triggering an earthquake at a low level of shear stress. Assuming that coseismic movements increase the porosity in the fault core, Sleep and Blanpied find a steady-state condition in which pore pressure drops just after the earthquake and the process repeats, with pore pressure fluctuating around the hydrostatic value in response to the seismic cycle. They then go on to show, for certain ranges of parameters for faults with thick cores, that the shear creep during the interseismic period could dominate fault motion, rendering the fault aseismic, as is the case for the 'creeping' section of the San Andreas fault in California.

A forthcoming paper by Chester, Evans and Biegel¹ makes this model all the more plausible. They arrive at essentially the same conceptual model from a detailed geological study of the material in the cores of several subsidiary faults of the San Andreas system. The key observation is that fault shear is concentrated within a ten-centimetre-thick ultracataclasite zone at the centre of the fault core. These ultracataclasite zones are composed primarily of multiply reworked vein material, testifying to repeated episodes of high-pore-pressure fluid injection. The observations also satisfy a major constraint on such models of weak faults: although some high-displacement faults such as the San Andreas are weak, faults with less displacement are not so weakened. For example, although the San Andreas seems to be weak (at least, this is what heat-flow data indicate), borehole stress measurements at Cajon Pass show that the Cleghorn fault, a minor left-lateral fault only four kilometres from the San

1. Begun, D. R. *Science* **257**, 1929–1933 (1992).
2. Ruvolo, M., Disotell, T. R., Allard, M. W., Brown W. M. & Honeycutt, R. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1570–1574 (1991).
3. Horai, S. et al. *J. molec. Evol.* **35**, 32–43 (1992).
4. Marks, J. Am. J. *phys. Anthropol.* **85**, 207–219 (1991).
5. Gibbons, A. *Science* **257**, 1864–1865 (1992).
6. Martin, L. & Andrews, P. *Courier Forschungs-Institut Senckenberg* **69**, 25–40 (1984).
7. De Bonis, L., Bouvrain, G., Geraads, D. & Koufos, G. *Nature* **345**, 712–715 (1990).
8. Shea, B. in *Orang-utan Biology* (ed. Schwartz, J. H.) 233–245 (Oxford University Press, 1988).
9. Pilbeam, D. R., Rose, M. D., Barry, J. C. & Ibrahim Shah, S. M. *Nature* **348**, 237–239 (1990).
10. Leakey, R. E., Leakey, M. G. & Walker, A. C. Am. J. *phys. Anthropol.* **76**, 289–307 (1988).
11. Diamond, J. *The Rise and Fall of the Third Chimpanzee* (Random House, New York, Hutchinson/Radiant, London, 1991).

Andreas, is strong, with a friction coefficient of at least 0.6 and normal pore pressure⁷. The observations of Chester *et al.* can explain this, because they show that the weakening results only from the presence of a well-developed fault core, which is itself the product of large fault displacement.

Hydrofracture constraint

There is, however, one major problem that remains to be resolved. It is thought that a brittle material can never resist a

with the fault without hydrofracturing occurring is 60°. Data on stresses in California are ambiguous on this point, suggesting that the maximum compression lies between 60° and 80° from the San Andreas fault^{1,8}. The lower end of this range is compatible with the hydrofracture constraint, but the upper end is not, even in the framework of Sleep and Blanpied's model. The problem is that it is not clear if the lower end will also be compatible with the degree of weakening implied by the heat-flow constraint.



Banning Fault, on the San Andreas fault system in southern California. The fault acts as a barrier to moisture, so that water running off the highlands in the background is trapped, producing a characteristic line of vegetation. (M. Rymer, US Geological Survey, Menlo Park.)

pore pressure in excess of the least compressive stress without drainage occurring by hydrofracturing. This limits the extent of weakening by the raising of pore pressure. Sleep and Blanpied, and Chester *et al.*, are each aware of this constraint, and argue that hydrofracturing will be suppressed by increased mean stress in the core of the fault, owing to its strength contrast with the surrounding rock.

But this misses the point: the problem still exists at the seal at the fault core/country rock interface. In the models, this seal is assumed to be impermeable, but is also tacitly assumed to be impenetrable to hydrofracturing. I cannot think of a plausible material with this property. For example, in the experiment of Blanpied *et al.*, the seal was country rock with pores cemented with deposited material. I imagine this material would still hydrofracture, but in their experiments they did not reach this condition and so did not test this possibility.

For an active strike-slip fault with friction coefficient 0.5, say, it can be readily shown that the largest angle the maximum compressive stress can make

These remarks also apply to the case of low-angle detachment faults in extensional regions, which have also recently been argued to be active as a result of high pore pressure⁹. There the hydrofracture constraint is far more severe, because for such faults to be active in the brittle field means that pore pressure must approach the vertical stress, which in this case is the maximum principal compression. □

Christopher H. Scholz is at the Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York 10964, USA.

1. Chester, F. M., Evans, J. P. & Biegel, R. L. *J. geophys. Res.* (in the press).
2. Scholz, C. H. *A. Rev. Earth planet. Sci.* **17**, 309–334 (1989).
3. Byerlee, J. D. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
4. Rice, J. R. in *Fault Mechanics and Transport Properties in Rocks* (eds Evans, B. & Wong, T. F.) 475–503 (Academic, London, 1992).
5. Sleep, N. & Blanpied, M. L. *Nature* **359**, 687–692 (1992).
6. Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Nature* **358**, 574–576 (1992).
7. Zoback, M. D. & Healy, J. H. *J. geophys. Res.* **97**, 5039–5057 (1992).
8. Hickman, S. H. *Rev. Geophys. (suppl.)* **29**, 759–775 (1991).
9. Axen, G. J. *J. geophys. Res.* **97**, 8979–8991 (1992).

DAEDALUS

Past fears

HALLUCINATIONS, says Daedalus, usually draw on the senses. Some do so directly, like faces seen in the fire, or (more extreme) the rococo sense-fantasies generated by LSD. Others use stored sense data, such as illusions of absent friends or dead relatives. But one form of hallucination seems independent of the senses: that ghastly withdrawal symptom of alcohol addiction, delirium tremens. The sufferer typically sees small, horrible creatures — snakes, spiders, toads or insects — attacking or crawling over him.

This syndrome raises obvious questions. Could an alcoholic who had never seen a snake, a spider or any such small creature, suffer from delirium tremens? If so, where would he get his mental images of them? Daedalus points out that many people are deeply and irrationally frightened of snakes and spiders. He reckons that we carry instinctive mental images of these creatures, implanted genetically at a time when our savage ancestors found them really dangerous. Delirium tremens activates these inherited images and the fear that goes with them.

So DREADCO clinicians are visiting alcohol detoxification centres worldwide, especially in regions free of creatures typical of the syndrome. (Ireland, for example, has a flourishing alcohol culture but no snakes.) They are giving each centre a special DREADCO computerized video system. It is a sort of real-time photofit display, encoding the visual elements of a wide range of possible and impossible small animals and insects. The user can construct and refine any animal image simply by manipulating a pair of joysticks. Only a very strong-minded alcoholic could work the display during an attack of delirium tremens, so as to describe the creatures that were tormenting him. But Daedalus hopes that many will be able to use it later, to reconstruct the creatures of their recent ordeal. Statistical averaging of many such images should reveal the form of the species that so terrified our savage ancestors.

Our distant origins will thus be clarified. If the images turn out to be those of species that flourished a few million years ago in, say, Africa, then an African genesis will be indicated for mankind. If the creatures are Asian, or Australian, or even Irish, the theorists will have to think again.

This subtle argument can be extended. The widespread dream of flying, for example, may be a genetic memory of swimming or floating, thus supporting the theory that we evolved from some water-dwelling primate. More controversially, it may link us to the pterodactyls.

David Jones

Clams before Columbus?

SIR — Zoologists and geologists agree that specimens of *Mya arenaria* (the American soft-shell clam) from the Holocene found in Europe first appeared in the sixteenth century, after the voyage of Columbus¹. But we have dated a sample from the Kattegat region on the east coast of the Skaw in northern Jutland, Denmark, that pre-dates Columbus's voyage. This result implies that contact between America and Europe existed before the sixteenth century.

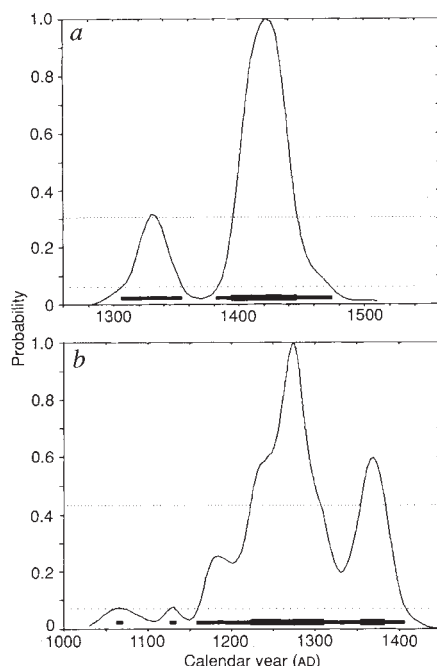
M. arenaria is known to have been transferred from America to Europe by man¹ because of zoological evidence that the larvae could not have spread spontaneously to Europe from America. The species has stayed in America since the Pliocene, whereas it died out in Europe at the beginning of the Pleistocene².

During our study of the Holocene coastal and faunal development of the Skagen Odde in northern Jutland, it became clear that the east coast offered the possibility of a close examination of the youngest fauna and coastal development³. For more than 1,000 years, the coast has migrated eastwards because of the accumulation of successive barrier sand systems.

The molluscan fauna of today is dominated by *Cardium*, *Spisula*, *Macoma* and *M. arenaria*. We took three samples from the east coast of the Skaw, all dominated by *Cardium edule* (common edible cockle), but with fragments of *M. arenaria*. The assemblage of molluscs from the site has been dated by conventional radiocarbon dating on the basis of *C. edule* using 50 g from each of the three samples. The *C. edule* shells were etched in hydrochloric acid to remove the surface layers. At least 10% of the mass of the shells was removed for analysis at the ¹⁴C dating laboratory in Copenhagen. The conventional radiocarbon datings of the three samples (K-5848 to K-5850) showed calibrated ages in the range AD 1400–1650. We subsequently radiocarbon dated one *M. arenaria* specimen from each of the three samples by accelerator mass spectrometry (AMS) using about 20 mg of each sample. The shells were given standard chemical treatment including acid etching to remove the surface layers (at least 20% of the total mass) for analysis at the AMS dating laboratory at the University of Aarhus. The age of the *M. arenaria* sample (II) found in the sand barrier furthest from the coast (AAR-883, AD 1245–1295 at ± 1 s.d.) without doubt predates Columbus's voyage in 1492. The figure shows the age probability distribution of the two radiocarbon datings of the oldest sample. It is obvious from the distribution that there is a

very slight probability of the sample being younger than Columbus's discovery of America in AD 1492.

These finds of *M. arenaria* raise the question of an earlier transfer from America to Europe than previously



The age probability distribution in calendar years for the two radiocarbon datings of the oldest sample (II). a, Conventional date on *C. edule* (500 $\pm$ 50 years BP). b, The AMS date on *M. arenaria* (720 $\pm$ 80 years BP).

assumed². In AD 1542, the French first colonized the east coast of America, but since the discovery of North America by Leif Ericson in around AD 1000, repeated trade by Nordic people settling in Greenland has been recorded. Consequently, *M. arenaria* could have been transferred from North America to Europe by the Vikings, but it still remains to be seen when *M. arenaria* invaded Europe.

K. S. Petersen

Geological Survey of Denmark,
DK-2400 Copenhagen NV, Denmark

K. L. Rasmussen*

¹⁴C Dating Laboratory,
National Museum and Geological Survey
of Denmark,
NY Vestergade 11,
DK-1471 Copenhagen, Denmark
J. Heinemeier, N. Rud
AMS Dating Laboratory,
Institute of Physics, University of Aarhus,
DK-8000 Aarhus C, Denmark

* Also at the Department of Physics, University of Odense, Campusvej 55, DK-5320 Odense, Denmark.

- Hessland, I. *Arkiv for Zool.* **37A**, No. 8, 1–51 (1945).
- Strauch, F. *Abh. senckenberg. naturforsch. Ges.* **531**, 1–210, (1972).
- Petersen, K. S. *Quatern. Int.* **9**, 53–60 (1991).

Temperature oscillations

SIR — The application of singular spectrum analysis to climatic temperature records has been the subject of recent debate^{1–4}. Tsonis and Elsner⁴ rightly emphasize the importance of considering several data records and illustrate the difficulty of determining whether or not two empirical orthogonal functions (EOFs) form an oscillatory pair by visual inspection.

Here, we resolve issues raised by the most serious response of Tsonis and Elsner⁴ to our note³. If a Toeplitz structure is imposed on the lag-covariance matrix, singular spectrum analysis gives two EOFs resembling interdecadal oscillations when applied only to the past 90 years of the IPCC global temperature series⁵. Tsonis and Elsner⁴ mistakenly attributed our remarks concerning this Toeplitz case to the corresponding non-Toeplitz case (shown in their figure), hence the disparity. In fact, the non-Toeplitz case they introduced illustrates an interesting special case of a degenerate singular spectrum, which we will discuss elsewhere. The fact that the Toeplitz approach of Ghil and Vautard seems to succeed better in this instance does not indicate conclusively that it is preferable, as the non-Toeplitz approach also has important advantages (such as unbiased reconstruction near the end-points of a linear trend; M.R.A. and L.A.S., manuscript in preparation).

In short, our aim³ was to explain why different conclusions from the IPCC data set were derived in refs 1 and 2. Our explanation gives no special role to the early decades; the different results arise from a simple artefact of singular spectrum analysis which does not establish (or require) the statistical significance of the oscillation. We suggest future work should avoid using rank order and visual appearance of EOFs as the primary significance criteria in singular spectrum analysis, focusing instead on testing quantitative null hypotheses.

Myles R. Allen, Peter L. Read

Department of Atmospheric, Oceanic and
Planetary Physics,
University of Oxford,
Oxford OX1 3PU, UK
Leonard A. Smith,
Mathematical Institute,
University of Oxford,
Oxford OX1 3LB, UK

- Ghil, M. & Vautard, R. *Nature* **350**, 324–327 (1991).
- Elsner, J. B. & Tsonis, A. A. *Nature* **353**, 551–553 (1991).
- Allen, M. R., Read, P. L. & Smith, L. A. *Nature* **355**, 686 (1992).
- Tsonis, A. A. & Elsner, J. B. *Nature* **356**, 751 (1992).
- Houghton, J. T. et al. *Climate Change, the IPCC Scientific Assessment* (Cambridge University Press, 1990).

A rete in the right whale

SIR — Whales maintain a homeothermic body temperature by losing heat to the surrounding water. The skin of the flippers, dorsal fin and tail flukes are the most effective areas used for heat dumping. The rest of the body is well insulated by subdermal blubber. Right whales (*Eubalaena glacialis*) have more of a problem maintaining a normal body temperature than other whales because they have a low surface area-to-volume ratio, extremely thick blubber, and no dorsal fin¹. The New England Aquarium, Boston, acquired a neonate right whale after it died (it had been found at Cumberland Island, Georgia, in January 1990). Dissection of the whale revealed a previously undescribed organ, an intraoral rete, that aids heat loss.

A rete, a body of convoluted blood vessels commonly found in mammals to store blood and regulate blood flow², connects arteries to veins, arteries to arteries and veins to veins³. Retial bodies were described first by Galen⁴. The intraoral rete of the right whale runs the full internal length of the upper jaw, with a total length of 83.8 cm in an animal 4.12 m long. It is an elongated 'Y' in shape, terminating in two large lobes. These lobes occupy a large gap in the baleen filter array.

Gross sections revealed a twisted mass of blood vessels, some muscle and no blubber. The rete is associated with a notch in the pterygoid plate of the skull. It connects with the basicranial rete, observed by Ommaney⁵ at the base of the skull external to the foramen magnum, allowing control of temperature for blood entering the brain. The brain would be protected from hyperthermic blood flow, an important mechanism because the brain is easily damaged by high temperatures.

Microscopy revealed a complex structure. The surface was covered with stratified squamous epithelium. Smooth muscle was present in three orientations from the maxilla: longitudinal groups appeared to run the length of the organ at the lateral extremity, and three to five bands of smooth muscle crossed the width of the rete in the horizontal plane. Small muscles radiated vertically from the jaw, extending from the longitudinal groups to the dermis. Very little areolar tissue was found.

Veins dominated the vascular structures. The lining of many veins resembled the transitional epithelium of a bladder. Cells were piled on each other in collapsed vessels that were clearly capable of much larger volumes. Arteries were clearly under-represented.

Small arterioles were found only in association with nerve bundles. Lymphatic vessels were seen scattered throughout the organ, but no lymphoid nodules were found.

A striking feature of the rete is a massive innervation. This is probably from the second division of the trigeminal nerve which serves the upper jaw in all mammals, and may be for vascular control or for temperature sensing. There was an abundant distribution of encapsulated mechanoreceptors at the dermis-epidermis boundary, possibly for temperature or hydrodynamic flow sensing.

The histological composition and anatomical position of the rete strongly suggest a thermoregulatory function. The organ functions by flushing its surface with cold water when the mouth is open, accomplished by relaxing the internal musculature and allowing the venous retial structure to flood with hyperthermic blood. As cold water flows over the uninsulated gingival surface, heat is readily lost.

Whales do not have sweat glands in the skin. The intraoral rete is an adaptive organ in balaenid whales that allows them to control their internal temperature by losing heat to oceanic water in various circumstances. We suggest that the intraoral rete may be present within and restricted to the Balaenidae (right, bowhead and pygmy right whales).

T. J. Ford Jr

Department of Biology,
Boston College,
Chestnut Hill,
Massachusetts 02167, USA

S. D. Kraus

New England Aquarium,
Boston,
Massachusetts 02110, USA

1. Omura, H. et al. *Sci. Rep. Whales Res. Inst.* **21** (Tokyo, 1969).
2. Ask-Upmark, E. *Acta Psychiat. Neurol. Suppl.* **VI** (1935).
3. Slijper, E. J. *Whales* (Cornell University Press, 1962).
4. Galen. *De Usu Partium Corpore Humanis Libris* (ed. Kuhn, P.) IX cap. IV, *Medicor Craecor. Opera* (Olms, Hildesheim, 1965).
5. Ommaney, F. D. *The Vascular Networks (Retia Mirabilia) of the Fin Whale (Balaenoptera Physalus)* (Discovery Reports, London, 1932).

Childhood thyroid cancer in Belarus

SIR — There is little doubt that the number of children reported to have thyroid cancer increased dramatically in radiation-contaminated areas of the Ukraine in 1990 (ref. 1) and in Belarus in 1990–91 (ref. 2).

It must be borne in mind, however, that intensive surveillance for thyroid cancer began in these areas in 1990 and that many of the cancers might perhaps never have been diagnosed otherwise. Histological examination of the thyroid from asymptomatic subjects is known to result in the detection of 'occult' papillary carcinomas which are indolent clinically and would probably never progress to cause frank symptomatic disease. At autopsy, the prevalence of occult papillary carcinoma is 2% at age 0–15 and 22% at age 16–30 (ref. 3). Most occult lesions found at autopsy with a diameter greater than 2 mm show invasion into the surrounding tissue⁴. Fewer than half the thyroid cancers which cause clinical disease are papillary⁵ and the fact that 128 out of 131 reported in Belarussian children were papillary carcinomas suggests that many, including those described as invasive, might be occult.

More information is needed on the way the thyroid cancers manifested themselves in children from Belarus, on the frequency of thyroid biopsies over time and in different regions, and on the geographical variation in reported thyroid cancer incidence. In particular, the following issues need to be addressed.

First, how many thyroid cancers in children manifested themselves with clinical

disease? How many cancers were picked up through investigation of asymptomatic thyroid enlargement? Has the number of thyroid biopsies in children increased over time and does this vary by region? Has the proportion of biopsies in which cancer was diagnosed changed over time and does it vary by region?

Second, what are the incidence rates of childhood thyroid cancer in each region? (It is unsatisfactory to report number of cases without referring to population size.) How do these relate to the overall pattern of radiation contamination in Belarus? If due to radiation, thyroid cancer incidence rates should be highest in the most contaminated districts (raions) of Gomel, that is, those which are very close to the Ukrainian border. Data on thyroid cancer incidence rates in each raion within Gomel are readily accessible and need to be reported.

Finally, an epidemic of thyroid cancer due to radiation exposure would be expected to continue over many years. It will be important to monitor trends in the future. The apparent increase in childhood thyroid cancer in the Ukraine and Belarus is of obvious concern, but it

1. Prisyazhiuk, A., Pjatak, O. A., Buzanov, V. A., Reeves, G. K. & Beral, V. *Lancet* **338**, 1334–1335 (1991).
2. Kazakov, V. S., Demidchik, E. P. & Astakhova, L. N. *Nature* **359**, 21–22 (1992).
3. Franssila, K. O. & Harach, H. R. *Cancer* **58**, 715–719 (1986).
4. Harach, H. R., Franssila, K. O. & Wasenius, V.-M. *Cancer* **56**, 531–538 (1985).
5. Pettersson, B., Adami, H.-O., Wilander, E. & Coleman, M. P. *Int. J. Cancer* **48**, 28–33 (1991).

is premature to attribute it entirely to the Chernobyl accident.

Valerie Beral

Gillian Reeves

*Imperial Cancer Research Fund,
Cancer Epidemiology Unit,
University of Oxford,
The Radcliffe Infirmary,
Oxford, OX2 6HE, UK*

SIR — V. S. Kazakov *et al.* and K. Baverstock *et al.* (*Nature* **359**, 21–22; 1992) report a sudden, unexpectedly large increase in the number of thyroid cancer cases in children in some areas of Belarus, starting about four years after the Chernobyl accident — considerably earlier than expected. The data provided in these reports are limited and preliminary in that they do not allow one to state whether the suggested increase in thyroid cancer cases is unequivocally attributable to radiation exposure. Many questions need to be answered, including the following.

Information on radiation dose (not reported by either Kazakov *et al.* or Baverstock *et al.*) is critical in determining the relationship to radiation exposure. In the case of children in Belarus, the focus of interest is the internal radiation exposure from radioactive iodine. However, several dosimetric problems remain. For example, it is difficult to examine the relative contribution of ¹³¹I and other short-lived radioiodines in exposed individuals. Also, uptake and elimination rates of iodine will be affected by iodine deficiency, which may occur in children from some of the affected areas, further complicating dose estimation. Efforts are now being undertaken by various groups to reconstruct the radiation doses of exposed individuals, including thyroid doses in children in the affected areas. It would be worthwhile to determine the thyroid cancer incidence in those children in Belarus for whom reliable thyroid doses are available.

Another question is whether the increased number of thyroid cancer cases reported in recent years really represents an increased incidence of this disease in a defined population. Efforts undertaken by expert pathologists to verify diagnoses is obviously an important first step in trying to answer this question. Detection of thyroid cancer is highly dependent on the intensity of medical screening because many of these cancers are clinically dormant and progress very slowly. Any attempt to estimate the true incidence of thyroid cancer after Chernobyl should carefully consider the potential effect of the increased general concern over thyroid cancer and the introduction of widespread medical screening for this cancer in the affected areas. The intensity of case ascertainment may also vary in different areas. It

is important to undertake a systematic ascertainment of cases, including thyroid tumours and other thyroid disorders. Kazakov *et al.* and Baverstock *et al.* seem to downplay the role of improved case ascertainment as the explanation for the increased number of recorded cases, but we do not know how many of the recorded cases were detected as a result of the medical screening and how many cases clinically manifested themselves. Finally, and most important, how will these numbers be expressed in terms of incidence rates in exposed and non-exposed populations?

We fully agree with Baverstock *et al.* that “understanding the consequences of Chernobyl will provide an important basis for preventive action in future”. However, we would add that studies of these consequences must be carefully pursued and based on scientific methodologies. One should not be overly alarmed, nor feel unjustifiably secure, on the basis of evidence that is not definitive. As in the past, our foundation is prepared to make its expertise, obtained in our long years of research on atomic-bomb survivors, available to assist in resolving some of the problems we have identified above.

I. Shigematsu

J. W. Thiessen

*Radiation Effects Research Foundation,
5-2 Hijiyama Park,
Minami-Ku,
Hiroshima, 732 Japan*

DNA fingerprints of cell lines

SIR — Few will disagree with the view expressed by researchers at the European Collection of Animal Cell Cultures (ECACC) that DNA profiling will transform authentication of cell cultures¹. But when? For the technique to fulfill its undoubted potential requires persuading originating investigators to publish DNA profiles of cell lines — itself contingent on standardization of probes and methods. Neither prospect is in sight. Hitherto, DNA profiling has been only fully effective in revealing those (presumably infrequent) contaminations occurring after accession of cell lines by cell banks themselves.

Although original cell lines will have highly specific DNA profiles, in the absence of any reference standards, these cannot be used *per se* to exclude contamination by unknowns. By contrast, isozyme typing (for quick and inexpensive species confirmation), cytogenetics (derided by the ECACC) or immunophenotyping (which they do not perform) are able to yield early clues to possible contamination. In this cell bank

we have been able to detect contaminated, wrongly identified or incorrectly characterized cell lines using some or all of these methods together with DNA profiling. Given the wide variety of cell lines held by large facilities, no single method could cope with the almost unlimited possibilities for contamination — a point emphasized by the US Foods and Drugs Administration in its guidelines for authenticating cell lines.

The ECACC's conclusion that cytogenetics lacks the capacity to provide “new approaches” for characterizing cell lines is open to question. The development of chromosome “painting” by fluorescent *in situ* hybridization allows for the first time the unambiguous identification of marker chromosomes and small unbalanced translocations (notorious bugbears of permanent cell lines) — and is potentially amenable to a far higher degree of automation than possible hitherto.

It remains to be seen whether the present domination by commercial interests of probe methodologies will promote or hinder standardization. Until then cell banks would be wise to hold on to their microscopes (and their DNA samples).

Roderick A. F. MacLeod

*Dept of Human and Animal Cell Cultures,
German Collection of Microorganisms and
Cell Cultures,
Mascheroder Weg 1B,
D-3300 Braunschweig, Germany*

STACEY *ET AL.* REPLY — MacLeod's comments, although valid in some circumstances, can only serve to mislead if, as in our article¹, the problem of identity testing the increasing diversity of cell cultures is to be addressed. Our conclusions were drawn from experience in a service department handling 1,100 different cell lines from at least 44 different animal species, which represents a wider range than that experienced in many laboratories. Few, if any, publicly funded culture collections have the necessary resources to carry out diverse and time-consuming cytogenetic techniques such as chromosome painting, fluorescence-activated chromosome sorting or analysis of interphase nuclei.

It is our experience that scientists need their cultures “this week” and although it is necessary to be pragmatic to meet this demand, it does not have to be at the expense of quality. We cannot attempt to duplicate depositors' research analyses; nevertheless we can maintain each culture stably and use a combination of isoenzyme analysis and DNA fingerprinting to identify cultures and exclude cross-contamination. Additional techniques (for example immunological analysis) are used when they are considered to be essential². Therefore the

recipients of cell lines can be confident in (and satisfied with) the material received, which is, after all, the point of the exercise.

Culture collections should provide the gold standard in authentication and are in the best position to cross-reference all the techniques used. Ideally, such an approach should be made by culture collections on a collaborative basis and lead to availability of identical material in cell banks as recommended by the World Health Organisation³.

In conclusion, the environment within which most culture collections operate demands that authentication procedures are not only accurate but also practicable. Although acknowledging that exhaustive characterization is a worthy goal, it would severely limit the range of material available. Most researchers

would not be able to afford the necessary fees to establish such a high level of quality assurance. Hold on to your microscopes for cytogenetics by all means, but remember that you will need to find the time and money to use them.

Glyn N. Stacey

Bryan J. Bolton

Alan Doyle

*European Collection of Animal Cell Cultures,
Public Health Laboratory Service,
Centre for Applied Microbiology &
Research,
Porton Down, Salisbury,
Wiltshire SP4 0JG, UK*

1. Stacey, G. N., Bolton, B. J. & Doyle, A. *Nature* **357**, 261–262 (1992).
2. Dewar, A. E. et al. *Biologics* **18**, 289–294 (1990).
3. World Health Organisation *Acceptability of Cell Substrates for Production of Biologicals* Tech. Rep. Ser. 747WHO (Geneva, 1987).

the 'ancestral' integument that characterizes all other unequivocal Devonian seeds. Earlier data suggested⁷ a major diversification of seed plants by the Late Tournaisian that included forms with winged seeds; this indicated adaptive radiation on a large scale, with implications of ecologically diverse strategies of pollination, dispersal and colonization^{8,9}.

The winged integument is interpreted as being an adaptation to aid dispersal. Enhancing the aerodynamic properties of the Devonian seed would boost the dispersal potential of the propagule by increasing flotation in the air and saltation along the ground. Besides improving dispersal, a winged integument also increases the probability of adherence to a substrate suitable for continued growth of the seedling by optimizing the shape and surface area of the seed. This viewpoint differs from previous interpretations of the functional adaptation of the integument, which have stressed the importance of protection¹ and optimization of pollination¹⁰.

A range of ecological strategies was developed by a rapidly diversifying group of seed plants well before the Late Tournaisian. This diversification is now documented in the form of a highly divergent, winged integument morphology before the end of the Devonian. It is, therefore, premature to consider integument function solely in terms of pollination and/or protection. Selective advantages among Devonian seed plants involved specialized dispersal mechanisms and these are now documented in the form of winged integument lobes.

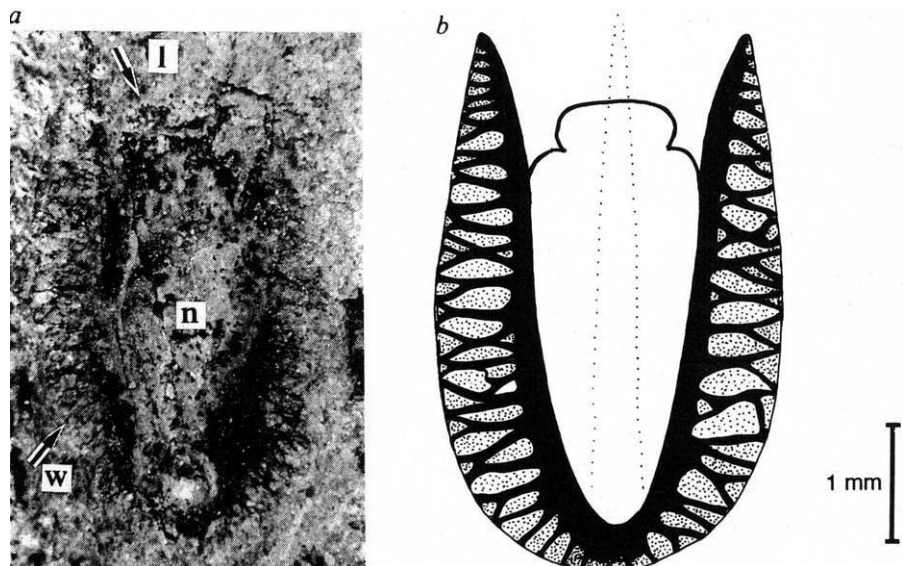
N. P. Rowe

*Department of Geology,
Royal Holloway and Bedford New College,
University of London,
Egham Hill,
Surrey TW20 0EX, UK*

Winged Late Devonian seeds

SIR — The evolution of the seed in 'ancestral' gymnosperm plants represented a major adaptive breakthrough in terms of pollination, protection and propagule dispersal¹. Seeds from Devonian gymnosperms show remarkably little variability in the structure and presumed function of the integument,

preted from critical observation and sections of fossils in the sediment. These are (1) the nucellar region (n); (2) the lagenostome (l) (a specialized funnel-shaped opening to allow entry of pollen); (3) winged integument lobes (w) consisting of inner, longitudinally aligned organic material, and a broad



a, Compression of seed showing evidence of two winged integument lobes; n, nucellar region; l, lagenostome; w, winged integument lobe. Magnification, 15 $\times$. **b**, Simplified diagram of seed compression surface with winged integument lobes. The dotted outline indicates the position of another integument lobe orientated perpendicularly to the plane of compression.

which consisted of simple narrow lobes^{2–5}. Late Devonian seeds with a specialized 'winged' integument discovered in central Germany⁶ represent the earliest evidence of an integument adapted for wind dispersal in seed plants.

The new fossil seeds have three main components (see **a** in figure) as inter-

outer region with conspicuous transverse 'fibres'. Up to four winged lobes have been observed in some seeds, with additional lobes visible as one or two longitudinal grooves perpendicular to the plane of compression (dashed outline in **b**).

This highly derived integument morphology is considerably different from

1. Rothwell, G. W. & Scheckler, S. E. in *Origin and Evolution of Gymnosperms* (ed. Beck, C. B.) 85–134 (Columbia University Press, New York, 1988).
2. Pettitt, J. M. & Beck, C. B. *Contr. Mus. Paleont. Univ. Mich.* **22**, 139–154 (1968).
3. Matten, L. C., Lacey, W. S. & Lucas, R. C. *Bot. J. Linn. Soc. Lond.* **8**, 249–278 (1980).
4. Gillespie, W. H., Rothwell, G. W. & Scheckler, S. E. *Nature* **293**, 462–464 (1981).
5. Fairon-Demaret, M. & Scheckler, S. E. *Bull. Inst. R. Sci. nat. Belg.* **57**, 183–199 (1987).
6. Higgs, K. & Streel, M. *Cour. Forsch. Senckenberg* **67**, 157–179 (1984).
7. Long, A. G. *Trans. R. Soc. Edinb.* **69**, 267–293 (1975).
8. Dimichele, W. A., Davis, J. I. & Olmstead, R. G. *Taxon* **38**, 1–11 (1989).
9. Galtier, J. & Rowe, N. P. *Nature* **340**, 225–227 (1989).
10. Niklas, K. J. *Bot. Rev.* **51**, 328–386 (1985).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Intricacies of nature

Colin Tudge

Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management. Edited by Peggy L. Fiedler and Subodh K. Jain. *Chapman and Hall*: 1992. Pp. 507. £52, \$79 (hbk); £22.95, \$35 (pbk).

Wildlife Research and Management in the National Parks. By R. Gerald Wright. *University of Illinois Press*: 1992. Pp. 224. \$32.50.

CONSERVATION, so long regarded as the natural preoccupation of woolly minds in woolly hats, has become the most intricate of disciplines: one in which the very best of science is still feeling its way, and in which the final arbiter, I believe, must be an emotional response that is essentially aesthetic. After all, the possibilities within any one landscape are effectively infinite, and what we choose to do will in the end be largely a matter of taste. The intricacy is revealed in these two excellent books.

What now complicates matters enormously is the end of the 'paradigm' of 'balance of nature', paradigm being the term very properly applied by Steward Pickett and others in *Conservation Biology*. Within that paradigm was the tenet of ecology that dates from F. E. Clements in 1916, and says that all ecosystems steer inexorably towards a 'climax' state which, once reached, effectively persists for ever. Of course, says Pickett, the notion of 'succession' is still extant, with one vegetation type tending always to give way to another. But the course the succession takes is now known to depend very much on circumstances: what was there in the first place, the order in which the various species invaded, and the occurrence and frequency of fires and diseases and so on along the way. Certainly, many vegetation types do have a 'climactic' feel to them, but these should be seen as special cases and as part of a dynamic system, not as an inevitable end-point.

Furthermore, the old balance-of-nature paradigm tended to assume that each ecosystem was self-contained and therefore self-regulating: indeed, that "any unit of nature is, in and of itself, conservable", and therefore that "any unit of a landscape can be an adequate nature reserve". But modern studies reveal the extent to which ecosystems depend on context: a stretch of forest surrounded by city is very different from the same stretch surrounded by more forest. The 'balances', such as they are, are certainly not entirely internal: any system can be swung this way and that by outside agencies. Instead of 'balance', says Pickett, we might speak of the 'flux' of nature.

This change of view has profound practical consequences, for the question

now arises: should we set out to conserve landscapes (habitats) or to perpetuate the processes that brought those landscapes into being? For example, Rutgers University in the 1950s set out to conserve Mettler's Wood, the last remaining uncut upland forest in central New Jersey. They protected the oak trees beautifully, restricting public access to a single trail with an expert guide. But the trees are now senescent. The forest is not regenerating. It is clear, now, that oak forest of this type does not regenerate unless there are regular fires — roughly every 10 years, to judge by the scars in ancient timber. Presumably, then, if we want the wood to continue for the next few thousand years, we should set light to it. That takes some courage plus a lot of (expensive) public education, among other things. But then

again, continuation in a pristine state may not be possible, for context again is all, and a fire in an isolated, remnant wood is very different from a fire in one patch of a vast forest.

On the world's naturally shifting landscapes we now superimpose mobile and all-consuming human beings. *Inter alia* and speaking broadly, as John Harper does in *Conservation Biology*, we are about to impose temperature rises on the entire planet higher "than have been experienced since the evolution of *Homo sapiens*". This in itself would make a nonsense of many of the present-day reserves that are the last outposts of wild creatures. We are finding it difficult to keep what we have: yet we may soon have to "provide the principles and technology to reassemble it, like London Bridge, elsewhere".

R. Gerald Wright, in *Wildlife Research and Management in the National Parks* (where 'the' means 'North American'), shows how complex the issues are even without the pending greenhouse effect. America's oldest national park, from 1872, is the mighty Yellowstone, while the National Park Service Act that created the National Park Service came on line in 1916. The act was clear enough in its aims — "to conserve the scenery and the natural and historic



Bear necessities — American hunters today are licensed to cull thousands of black and brown bears annually. The bears also clash with foresters, bee-keepers and picnickers. Problems with picnickers began in the 1930s, when bears feeding on refuse in national parks became a major attraction. After much havoc and controversy, park refuse dumps were fenced in 1971 and hundreds of bears starved, with others roaming in search for food outside the parks. In 1972, there was the first fatal mauling of a person in the United States by a bear in 30 years. This picture of an American black bear is taken from *The Velvet Claw* by David Macdonald, which is published to accompany a BBC television series on the history of carnivores (BBC Books, £17.95).

objects and the wildlife therein and provide for the enjoyment of the same" — but the scope it leaves is enormous, neglecting such aspects as whether to conserve landscape or species, *status quo* or the processes that created the *status quo*; to what extent to restore what was (for example the wolves to Yellowstone); and to what extent to gear the parks to human enjoyment, which meets the terms of the act and helps to pay the rent, but also — given the present state of the art — compromises the wildlife. Today, although all working within the act, different parks pursue very different policies.

In practice, too, conservationists are not the only people with a say. Farmers are involved: for instance, attempts in the 1980s to conserve the grizzly bear, which is threatened in the 48 contiguous US states, provoked the National Wool Growers' Association to opine that "the loss of one human life is not worth the entire grizzly population". In practice, of course, the number of deaths due to grizzlies in national parks — one per 30 million visitors — is many times lower than those caused by drowning, motor accidents and other human follies. Then again, foresters have a very different attitude to fire control than do modern conservation biologists. And so on.

Science cannot of course solve the political disputes, although it can help to predict the outcome of any particular policy. Whether the scientists are listened to, however, or even enabled to find out the facts in the first place, is a very different matter. National Park Service scientists are doing much excellent work, not least in monitoring, and the observations of rangers are just beginning to be fully appreciated. But, says Wright, there was no national science programme in the parks until the 1960s; only 1.5 per cent of the total budget of the National Park Service is spent on research; the research stations that do exist — Great Smoky Mountains, Everglades, Yosemite — are far from the parks' administrative centres; and no natural scientist has ever been in charge of the agency, while "few have even been regional directors".

In short, we cannot possibly pick our way through the options, or even see what the options are, without good science. But the fact that science even has a serious role to play is not yet fully recognized by the people who are really in charge.

These books approach essentially the same problems from very different angles. Both are valuable contributions, and deserve to be bought and read. □

Colin Tudge is a zoologist and science writer at 208 Clive Road, London SE21 8BS, UK.

Planetary perception

David W. Hughes

The Planets: Portraits of New Worlds.

By Nigel Henbest. Viking: 1992. Pp. 207. £20, \$35. (US publication in December.)

Worlds in the Sky: Planetary Discovery from Earliest Times Through Voyager and Magellan.

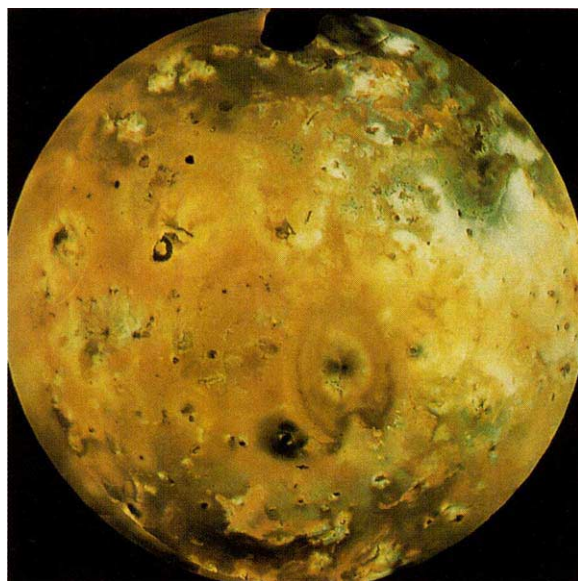
By William Sheehan. University of Arizona Press: 1992. Pp. 243. \$35 (hbk), \$17.95 (pbk). (UK distribution by Roundhouse.)

THE logarithmic sequence 3,000, 300 and 30 simplistically represents the duration of the three ages of planetary exploration. The first age was that of the naked eye. Planets were seen as no more than bright star-like points. But coupled with

into three phases. Scientists have nearly finished the first, this being the one in which spacecraft hurtle past the planets taking as many measurements as they can in the brief time they are close. Henbest concentrates on this phase because, for most of the planets, the second phase, in which spacecraft are placed in orbit, is yet to come. Two exceptions are the Magellan spacecraft in orbit around Venus and the Viking spacecraft around Mars, both of which have already produced a huge amount of data. The next decade will see the Galileo and the Cassini probes going into orbit around Jupiter and Saturn. Henbest bypasses the phase where robot probes will return with samples from the planets and looks forward to the third phase, which although "still some way off . . . sees the exploration of our fellow worlds by human beings".

The Planets is lavishly illustrated in colour and the pictures are of the highest

From *The Planets*



Jupiter's moon Io, racked by immense sulphurous eruptions, as viewed by the Voyager spacecraft.

the normal diurnal movement of the sky, the planets wandered along the ecliptic band with variable speeds complicated by a combination of prograde and retrograde motion. The problem of explaining this motion was sufficiently difficult to keep some of the best astronomical minds hard at work. Thankfully, it was not so difficult that they gave up. It was eventually solved in the seventeenth century by Kepler and Newton.

The middle period is the age of the planetary telescope. Here we see the discovery of three new planets, multiple satellite systems, mountains, craters, rings, red spots and even canals. The final period, an age that has already lasted 30 years, is the space age.

Nigel Henbest concentrates solely on this latter period. Following the tripartite approach, he divides the space age

quality I have seen. In fact, 70 per cent of the book is pictures. In the accompanying nontechnical text, Henbest has done a first-class job in conveying the excitement of modern planetary exploration. Each newly explored planet seems to reveal information useful to Earth dwellers. On Venus we see the alarming results of a greenhouse effect that has got out of control. Mars suffers from the ravishes of a perpetual ice age. Jupiter, Saturn, Uranus and Neptune have atmospheres that account for most of their mass and they also have vastly differing solar energy inputs. This makes them perfect laboratories for studying weather systems.

Interplanetary comparisons may even lead to better weather predictions on Earth. *The Planets* is an excellent, up-to-date, general introduction to space-age planetary astronomy.

William Sheehan has written an entirely different type of book, albeit on the same subject. He concentrates on the telescopic phase of planetary exploration. Many amateur astronomers have given up telescopic observations of the planets in the simple belief that the space probe can do it so much better. But not Sheehan. His text conveys a detailed understanding of the historical nuances of planetary discovery coupled with a great love of gazing through telescopes. To quote the author and H. P. Wilkins, "the actual sight of lunar mountains or of planetary markings imparts a sense of satisfaction which the

armchair astronomer neither knows nor deserves to know".

Sheehan's aim is to place the heroic work of the past in the context of recent discoveries. In doing this he delicately juxtaposes the maps of Mercury made by Giovanni Schiaparelli (in 1889), Eugène Antoniadi (in 1933) and Audoin Dollfus (in 1972) with the images returned by the Mariner 10 spacecraft. Percival Lowell's 1895 descriptions of Mars are compared with the Viking Orbiter and Mariner 9 images. Harold Hill's pen-and-ink drawings of lunar features are placed against the Apollo photographs. Sheehan provides intricate descriptions of such topics as the discovery of the asteroids, the quest to understand the physical form and origin of the rings of

Saturn, the use of the perturbation of the orbit of Uranus to predict the position of Neptune and the search for Pluto. Like Henbest's book, Sheehan's account is nontechnical. In the historical context this means that footnotes do not abound and that mainly secondary sources are used.

Worlds in the Sky captures the excitement of the astronomical quest for planetary detail and understanding. By carefully tracing the steps that have been made in the past, Sheehan makes us even keener to draw back more veils in the future. □

David W. Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

Science and salvation

H. N. V. Temperley

Science and Religion: Some Historical Perspectives. By John Hedley Brooke. Cambridge University Press: 1991. Pp. 422. £27.50, \$44.50 (hbk); £10.95, \$12.95 (pbk).

THIS is an astonishing book about one of the most important problems facing our culture. Down the ages, the relationship of science to philosophy and religion has changed in countless ways. The main theme of the book is the almost unbelievable subtlety, complexity and diversity of this relationship. One cannot help but admire the author's vast reading, his penetrating critical power, his grasp of detail and his ability to summarize.

The three traditional views, that of conflict, that of complementarity and the intermediate view that science and religion have much in common, are all seen to be oversimplifications. In a chapter on natural theology, for example, Brooke shows how the apparently simple issue raised by Paley of whether the Universe shows evidence of conscious design in the same way that examination of a watch does started a huge argument that still rages today. Brooke describes how thinkers quickly became involved in the problems of evil, the subtle discussions of Hume and Kant of the notions of cause and effect, and the unfinished argument of how far darwinian natural selection can account for the development and changes of animals and plants.

Paley would, to my mind, have done better to compare the Universe not to a finished product such as a watch but to a factory or nursery garden where there is not only continual development and change but also losses, mistakes and failures. In this scheme, creation would be a continuing process rather than a

single act in the past. The subsequent course of the argument would probably have been similar to, but perhaps even more complicated than, that described.

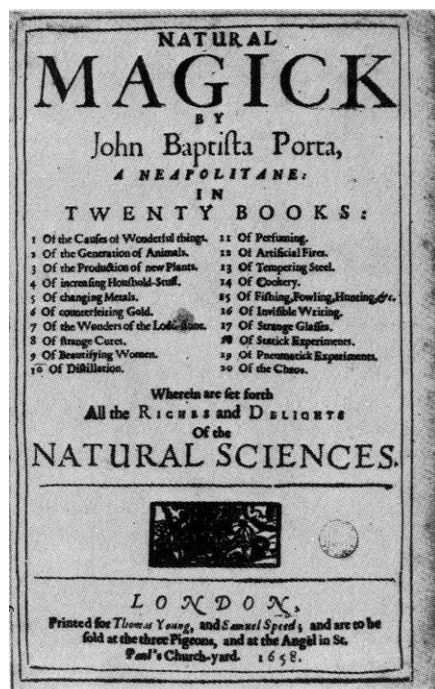
Following an introduction, summary and some preliminary considerations, Brooke gives an account (Chapter 2) of the scientific revolution and the emergence of science in the sixteenth and seventeenth centuries from alchemy and magic. Religion was developing rapidly at the same time and we get many fascinating glimpses of the resulting interactions and differentiations. This chapter is packed with enough information to provide material for several

books and ends with accounts of Boyle, Galileo and Halley. In the next chapter, Brooke investigates whether parallels can be drawn between the Copernican revolution and the Protestant Reformation, but he is really asking whether reform of religion helped the expansion of science. He warns of the difficulties of testing generalizations about the tolerance of Protestant societies towards new scientific learning, and is cautious about coming to any conclusion.

Chapter 4 addresses the question, first arising in the seventeenth century but still with us, of what room there is for divine intervention in a Universe governed by scientific laws. In the seventeenth century, the clockwork image of nature was enlisted to reinforce the view that nature was a designed system, affirming God's absolute sovereignty. There is an account of some of the struggles that Descartes, Boyle and Newton had reconciling these stubborn problems, in particular Newton's notions of "ordinary" and "extraordinary" providence and his attempts to reduce the Bible to order in the same way that he had reduced nature to a collection of laws. He set up a system of biblical interpretation that contained 15 different principles (evil spirits became mental disorders, devils became the imaginary ghosts of the departed, and so on). Perhaps he himself realized that this was becoming too complicated to be credible.

It is ironic that the same mechanical model of the Universe was used by the deists of the eighteenth century to attack established religion, one of the topics of the next two chapters, "Science and Religion in the Enlightenment" and "The Fortunes and Functions of Natural Theology". The newly emerged sciences were trying to find their place in society; hailed as instruments of progress and in an age when unprecedented confidence was placed in the power of human reason, they were soon transformed into a secular force. Brooke describes their impact on politics and, in Chapter 7 ("Visions of the Past"), their influence on studies of human history and the history of the Earth. During this time, the relationship between religion and science vacillated between open conflict and attempts at reconciliation.

It is in Chapter 8 on the darwinian challenge to popular Christianity that Brooke specifically tackles the difficult question of the extent to which the evolutionary forces of natural selection can mimic the creative powers of God. A common view has been that evolution can explain many of the steps in life's long journey from single-celled organisms to humans, but not certain aspects of it, such as the development of the eye and ear. Equally hard to explain, so it was held, are the balance and coordina-



Title page of J. B. Della Porta's *Natural Magick*, first published in Latin in 1558.

tion of functions of the body. The argument was that these organs and functions have to be nearly perfect or else they are worse than useless to the organism. However, this view is now challenged by the majority of evolutionary biologists, perhaps most forcefully by Richard Dawkins, who in *The Blind Watchmaker* (Longman, 1986) claims that evolutionary change can explain all.

The final chapter, modestly described by the author as a "postscript", deals with a few twentieth-century developments such as Freud's 'explanations' of religion, the revolution in physics and

the resurgence of the interest in science and human values. Last, but not least, is the "Bibliographic Essay" which fills more than 50 pages.

Brooke's vast reading and the impressive use he has made of it attest to the complexity of the relationship between science and religion. To use his words, there is no such thing as *the* relationship, whether it is couched in terms of war or of peace. □

H. N. V. Temperley is at Thorney House, Thorney, Langport, Somerset TA10 0DW, UK.

Preparing for the brain decade

Michael R. Hanley

Neurons and Networks: An Introduction to Neuroscience. By John E. Dowling. *Belknap (Harvard University Press): 1992. Pp. 447. \$45, £29.95.*

The Neuron: Cell and Molecular Biology. By Irwin B. Levitan and Leonard K. Kaczmarek. *Oxford University Press: 1991. Pp. 450. £55, \$65 (hbk); £18.95, \$39.95 (pbk).*

An Introduction to Molecular Neurobiology. Edited by Zach W. Hall. *W. H. Freeman (UK)/Sinauer (US): 1992. Pp. 555. £32.95, \$46.95.*

THE phrase 'coming of age' is used in two of the three introductions to these new neurobiology textbooks in recognition that the discipline has, in many senses, grown into adulthood after a protracted adolescence. Firm foundations of hard facts now exist in specialist fields of neurobiology, such as the molecular mechanisms of neural development where previously there were black boxes. Moreover, technical innovation, such as patch clamping and hybridoma technology, has made accessible previously intractable problems, such as the structural analysis of excitability proteins. It now seems that there is a cohesive core of knowledge about techniques and principles of neurobiology; and each of these texts takes a different aspect of this knowledge as its starting point. All three approaches work well, and the textbooks are uniformly excellent.

Neurons and Networks covers the classic material of neurobiology. It has evolved from years of lectures given to undergraduates at Harvard University by the renowned retinal physiologist John Dowling. The text is divided into two sections, "Cellular Neuroscience" and "Integrative Neuroscience", but the author has wisely focused on his own specialist area, the visual system, to develop concepts throughout the book.

Parts of the first section are weak. Specifically, the chapter on the chemistry of synaptic transmission is not as up to date as the corresponding material in the other two books. A telling gaffe is the repeated use of "inositol triphosphate" — a nonexistent compound — for inositol trisphosphate (my italics). The second section on integrative neuroscience is built on an evolutionary progression from invertebrates to vertebrates to higher cognitive functions, with an intervening chapter on developmental neurobiology. Overall, this is a sober and scholarly volume that strikes me, with its terse, no-frills style and reliance on the history of ideas, as very Harvard. Although intended as an introduction to basic neurobiology, it is a satisfying book at several levels.

The Neuron is a more advanced presentation, focusing on cell and molecular neurobiology at the level of graduate students. The authors, Irwin Levitan and Leonard Kaczmarek, are distinguished neuroscientists with a history of various types of collaborations in teaching, writing and research. In writing together, they have evolved a distinctive style, which is often slyly irreverent. The text is constructed from three sections: "Electrical Properties of Neurons", "Intercellular Communication" and "Behavior and Plasticity". The text is impressively modern, with up-to-date information on the trendiest areas of neurobiology, such as adhesion molecules, synaptic vesicle biogenesis and function, and ion channel systems. Much of this new material is unavailable in comparable texts. The book is highly visual, with figures on virtually every page. The figures deserve special comment because they are a teacher's dream: simple and uncluttered, but conceptually powerful. Frankly, although the recommendation is often abused, *The Neuron* is one of those books that really do belong on every shelf.

Introduction to Molecular Neurobiology is a multi-author volume edited by another distinguished neurobiologist, Zach Hall, and draws heavily on the editorial ranks of the journal *Neuron*. It is the most detailed and sophisticated of the three books. The chapters are organized as four sections: "Signalling in the Nervous System", "Cell Biology of Neurons and Glia", "Neuronal Development" and "Complex Interactions of Neurons and Neuronal Disorders". Because of its explicit focus on molecular aspects, this text includes extensive treatment of topics not found in the other two volumes, such as the molecular basis of developmental neurobiology and the genetics of neurological disease. The book is rich with figures and often uses inserts to introduce a self-contained discussion of a single subject, such as video microscopy, at a point where the subject is relevant but distracting if presented in the main text. Indeed, the visual presentation is so good that the text almost seems to be a giant figure legend. The book provides the comprehensive treatment of molecular neurobiology that has been sorely needed. I regard it as the definitive specialist text of the subject, and it is likely to remain so for some time, given the high standard of the contributions.

Lastly, there are important similarities between the texts which contribute to their outstanding quality as educational aids. They all present a summary at the end of each chapter, which restates the important generalizations that emerge from the details of specific experiments. Each book also recommends a few key references, emphasizing recent reviews and pivotal papers.

The expansion of neurobiology has led to greater integration of molecular, physiological, anatomical and systems studies than ever before, but has also created an unprecedented volume of information. The complete neurobiologist has to be familiar with a huge diversity of subjects from genetics to psychophysics. So is it wise, or even possible, to cover the spectrum of subjects now subsumed by neurobiology in a single volume? These books have not attempted to do so; rather, they have each defined a specific orientation that cuts across the main issues in neurobiology. They are not in competition, but complement each other, and all could be used effectively in a modern neurobiology curriculum. And researchers should consider adding them to personal libraries, as they mark the emergence of a new generation of modern neuroscience texts. □

Michael R. Hanley is in the Department of Biological Chemistry, School of Medicine, University of California, Davis, California 95616, USA.

Creep, compaction and the weak rheology of major faults

Norman H. Sleep* & Michael L. Blanpied†

* Department of Geophysics, Stanford University, Stanford, California 94305, USA

† United States Geological Survey, Mail Stop 977, Menlo Park, California 94025, USA

Field and laboratory observations suggest that the porosity within fault zones varies over earthquake cycles so that fluid pressure is in long-term equilibrium with hydrostatic fluid pressure in the country rock. Between earthquakes, ductile creep compacts the fault zone, increasing fluid pressure, and finally allowing frictional failure at relatively low shear stress. Earthquake faulting restores porosity and decreases fluid pressure to below hydrostatic. This mechanism may explain why major faults, such as the San Andreas system, are weak.

THERE is now ample evidence that many continental and oceanic strike-slip fault zones are weak^{1,2}. First, fault zones are weak relative to the surrounding country rock. Frictional failure involves traction (resolved stress on planes) and is expected where the ratio of shear traction to normal traction is high—that is, where the principal axis of maximum compression is ~60° from normal to the fault plane. Yet the plane of the San Andreas fault is unfavourably oriented, being nearly perpendicular to the regional principal axis of compression¹. Second, fault zones are weak in an absolute sense compared with laboratory experiments. For example, the absence of elevated heat flow over the San Andreas fault indicates that sliding causes little frictional heat and therefore that the average shear traction on the fault is less than 20 MPa (see for example refs 2, 3). At seismogenic depths on the San Andreas fault, ~7 to ~20 km (refs 4, 5), normal tractions across the fault plane are inferred to be ~200–570 MPa. For a representative laboratory value⁶ of 0.7 for the coefficient of friction, and assuming ‘hydrostatic’ fluid pressure (that is, in equilibrium with a column of water to the surface), the shear traction required for slip is 90–260 MPa, far higher than the 20 MPa inferred from heat flow.

Here we present laboratory experiments and physical modelling which imply that transient high fluid pressures and earthquake failure at low shear tractions are expected consequences of small amounts of ductile creep in fault zones. We construct a simple physical model of earthquake faulting which incorporates laboratory, field and geophysical evidence. Before presenting the model and results, we briefly review published explanations for why faults are weak, and summarize observations from laboratory and field studies on the role of fluids in the earthquake process.

Mechanisms for weakening fault zones

When discussing why the fault is weak, it is convenient to consider frictional failure on planar surfaces

$$\tau_f = F(T_n - P_f) \quad (1)$$

where τ_f is the shear traction at failure, F is the coefficient of friction, T_n is the normal traction on the fault and P_f is fluid pressure. Low normal traction T_n on the San Andreas fault is easily eliminated because the axis of maximum compression is nearly normal to the fault¹, implying that T_n exceeds the pressure caused by the weight of the overlying rocks. Thus, deep fault zones may be weak because the coefficient of friction F is unusually low, the fluid pressure P_f is permanently much higher than hydrostatic, the fluid pressure is transiently increased at the time of earthquakes, or the coefficient of friction is transiently decreased at the time of earthquakes. We do not discuss fault zones with very low intrinsic coefficients of friction because

laboratory measurements have not found plausible Earth materials (see, for example, ref. 7).

Permanently high fluid pressures. Fluid pressures approach lithostatic (rock) pressures in various geological environments. Low-grade and medium-grade metamorphic rocks are common examples. During prograde metamorphism the original pore water and water dehydrated from minerals are driven off faster than they can be removed by porous flow. The fluid pressure thus reaches the least compressive stress. In such circumstances, hydrofracture on planes perpendicular to the least compressive stress can coexist with pressure solution on planes perpendicular to the most compressive stress⁸. This mechanism occurs at subduction zones that are continually underplated by trench sediments⁹ as well as areas of basin formation, crustal thickening, and regional metamorphism.

Regionally high fluid pressures do not provide a good explanation for the observed orientation of the San Andreas fault along a plane of low shear traction, as more favourably oriented planes in the country rock would also be weakened by fluid pressure. One possibility is that the fault zone is so impermeable that leakage does not relieve high fluid pressures within the fault zone or increase them in the country rock over geological time¹⁰. The cause of high fluid pressure still needs to be explained, as many hard rock types may have little water initially. Surface water can flow down only if the pressure is less than hydrostatic and will leak back towards the surface if pressure is greater than hydrostatic¹¹. For example, a process of two or more stages may produce high fluid pressure in an initially dry rock: (1) cracks open and fill with surface water; (2) the cracks are isolated from the surface and are compacted by deformation of the rock until elevated fluid pressures occur. This sequence is inferred to have occurred in oceanic gabbros, from chemistry and cross-cutting relationships¹². Alternatively, a source of high-pressure water may exist below the seismogenic zone. The fault zone then acts as a conduit for water supplied from below and as an impermeable barrier to lateral flow into the country rocks¹³. A deep water source is unlikely for many faults with no underlying sediments, such as those in the French Pyrenees¹¹.

Transiently weak fault zones. The difficulties of maintaining high pore pressure in fault zones for geological periods of time without also pressurizing the country rocks are eliminated if the elevated pressure exists only for the few seconds of the earthquake or for several years in the earthquake cycle. Compaction of relatively porous fault zones by shaking during earthquakes, termed acoustic fluidization, is a possible mechanism¹⁴. Frictional heating may also increase the pressure in the pore fluid through thermal expansion¹⁵ or the dehydration of clays¹⁶, or may melt the rock along the fault zone^{15,17}. These mechanisms require considerable fault slip to operate and thus

TABLE 1 Fault model parameters

Constant	Value	Constraint
Depth, z	14 km	Field
Temperature, θ	300 °C	Field
Thermal gradient, θ/T_n	0.75 °C MPa ⁻¹	Field
Hydrostatic gradient, P_f/T_n	0.35	Field
Coefficient of friction, F	0.7	Lab
Slip rate, V_f	0.05 m yr ⁻¹	Field
Shear modulus, μ	3.5×10^4 MPa	Field
Elastic length scale, L	8 km	Field
Stress drop, γ	30%	Field
Maximum pore porosity, ϕ_p	0.07	Poor
Maximum crack porosity, ϕ_c	0.02	Poor
Maximum aspect ratio, $\alpha_{\max}$	120	Poor
Energy pore porosity, β_o	0.02	Poor

are probably ways that allow a magnitude 5 earthquake to grow into a magnitude 8, rather than mechanisms that initiate rupture.

Role of fluids in the earthquake process

We present a mechanism for initiating rupture involving two features of ductile creep that have been discussed since 1925 (ref. 18). Under appropriate conditions, ductile creep occurs at stresses far below those for frictional failure, and it leads to compaction whereas frictional failure generates porosity in a low-porosity rock. Specifically, we suggest that ductile creep leads to compaction so that fluid pressure increases gradually from below hydrostatic following an earthquake to a near-lithospheric value which permits the frictional sliding of the next earthquake. Laboratory and field evidence indicates that the mechanism may actually occur.

Evidence from laboratory studies. Laboratory experiments at high temperatures suggest that the behaviour of fault zones is strongly modified by chemical reactions of water with rock^{19,20}. Many brittle and ductile deformation mechanisms can be activated or accelerated in the presence of aqueous fluids, even at modest temperatures. Because these mechanisms are thermally activated, their relative importance is increased by higher temperatures and lower strain rates. Solution transport, the movement of dissolved constituents through the pore fluid, can reduce the intrinsic frictional strength of faults by allowing deformation mechanisms such as dissolution-precipitation creep or pressure solution^{20,21} which operate at low differential stress. Fluid permeability can be reduced markedly by solution-transport-aided compaction in porous aggregates²², and by healing of micro-cracks²³ and sealing of fluid pathways by precipitation²⁴.

Of particular relevance here are frictional sliding experiments²⁵ in which a self-sealing process occurred very rapidly, causing the formation of a low-permeability fluid barrier, or

pressure seal. This seal prevented the communication of fluids between a porous, shearing fault zone and the surrounding rock, which lead to overpressure and low sliding strength. The experiments were done on 'faults' in granite consisting of a sawcut surface filled with a 0.5-mm-thick layer of simulated fault gouge (granite powder). 'Pore water' was supplied at 100 MPa through a bore hole in the granite ending 3 mm from the fault. Samples deformed at 600 °C slid stably with low strength and an apparent coefficient of friction of 0.22, well below the intrinsic friction of granite at these conditions, roughly 0.5 (which is itself lower than the coefficient of friction for dry granite, ≥ 0.7). Tests with well-drained samples showed that the low strength resulted from sealing of the granite bounding the fault, causing water to be trapped at high pressure within the compacting, shearing fault gouge. A simple effective stress calculation demonstrates that the fluid pressure was about a factor of 3 higher within the fault than in the nearby bore hole and country rock.

Both the formation of the seal and the elevation of fluid pressure in these experiments are believed to be caused by compaction. The country rock probably became sealed because of precipitation from solute-laden water driven out of the compacting gouge layer. Continued compaction after the seal was in place, by solution transport and by fracturing and rearrangement of gouge particles^{26,27}, raised the fluid pressure in the sealed fault.

These experiments illustrate three processes relevant to faults in the Earth. First, the presence of water at high temperature lowers the intrinsic rock strength at low strain rates. Second, pore fluids in faults may become isolated from those in the surrounding rock by the formation of low-permeability seals. Although the experiments described above were done at 600 °C, seals also form at much lower temperatures in a variety of rock types^{24,28}. Given the years or decades typical of earthquake recurrence intervals, similar seals could probably form around natural faults at temperatures of 200–400 °C (ref. 29), more typical of seismogenic depths. Third, combined shear and compaction in the fault zone can cause an increase in pore pressure and allow sliding at low effective stress. Below, we will use these three observations to construct a fault zone model in which interseismic fault creep causes changes in pore volume and fluid pressure within a partially sealed fault.

Field evidence from exhumed faults. Detailed petrological studies of exhumed fault zones provide insight into the deformation processes that accommodate fault slip and the roles of pore fluid in the earthquake cycle. Rocks in seismogenic faults are alternately subjected to the high strain rates of the earthquake failure and the low strain rates of the interseismic period. Evidence for this alternation is given in exhumed fault zones by the mutual overprinting of fractures, friction melts and explosion breccias indicative of seismic slip, and pressure solution fibres, preferred crystallographic orientations and mineral

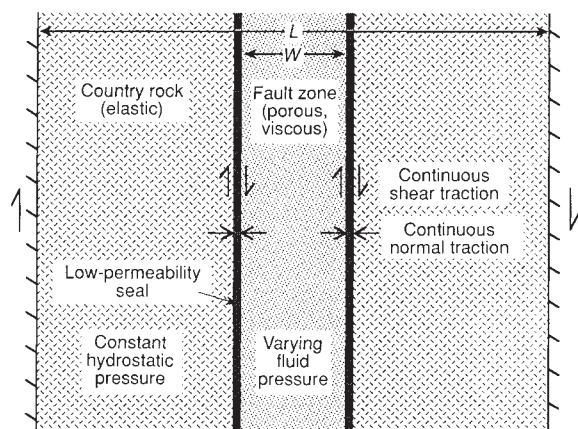


FIG. 1 Boundary conditions between the tabular one-dimensional fault zone and the country rock are shown schematically. L is the effective width of the region which deforms elastically. W is the width of the ductile fault zone. A narrow seal allows only low rates of fluid flow between the fault and country rock.

TABLE 2 Input parameters

Model	Zone width, W (m)	Intrinsic viscosity, η_i (yr)	Time step (yr)	Leak time, t_h (yr)	Energy crack porosity, β_c
1	1	10^6	2	1,000	0.02
2	10	10^5	2	1,000	0.02
3	100	10^4	0.3*	333*	0*

* Changed to aid numerical calculation.

reactions indicative of slow creep^{30–32}.

Evidence for ductile, solution-controlled creep combined with elevated pore fluid pressures is seen at many localities. Fine-grained quartz was deformed by pressure solution at palaeodepths of less than 2 km on the Dixie Valley normal fault in Nevada, USA³². Ductile creep and elevated fluid pressures are inferred to have occurred at a palaeodepth of 11 km on the Wasatch normal fault in Utah³³. Creep within fault zones is more rapid than in the country rock, because the faulting process reduces the grain size and alters the mineralogy of the fault towards micas, clays and hydrothermal minerals (such as quartz and calcite) which deform easily³⁴.

Branches of the San Andreas fault system exhumed from 2 to 5 km depth show evidence for low-permeability seals, elevated fluid pressure and sliding at low effective stress³⁵. Arrays of subsidiary faults in the surrounding rocks imply near-fault-normal compression, and therefore low sliding resistance. Rocks bounding these faults show many episodes of formation and healing of fractures; this indicates both that fluid pressures reached lithostatic levels at least during a portion of the earthquake cycle, and that permeability reduction processes are effective even at these fairly shallow crustal levels. Chester *et al.*³⁵ conclude that elevated fluid pressure was generated and maintained by a combination of factors, including episodic self-sealing of the fault-bounding rocks and time-dependent compaction within the fault zone. Coseismic dilatancy in the core of the fault is indicated by 'jigsaw puzzle' brecciation³⁶ and abundant hydrothermal veining. These features are consistent with isotope studies^{11,37} which require large volumes of metasomatizing fluids to have passed through many fault zones.

Compaction model for fault zones

We present a conceptual model for a strike-slip fault zone because the geometry is simple and many data are available for the San Andreas system. The key aspects have been discussed in the geological literature; in particular, the fact that faults are often bounded by low-permeability rock^{10,13,35,38,39}, that pressure-solution creep causes compaction and pore pressure increase^{35,40}, and that intraseismic creep and seismic porosity production occur within fault zones^{32,35}. To develop a coherent theory, we obtain a one-dimensional mathematical and numerical model following the methods and notation of ref. 41 and using basic mechanics given in ref. 42. Various heterogeneous two- and three-dimensional processes, including fault behaviour with depth, fluid flow within the fault plane^{11,13,38,43} and after-shocks, and detailed application of the heat-flow constraint^{2,3}, are beyond the scope of this paper.

Fault zone rheology. Following Rice¹³, we envisage a weak tabular fault zone in an elastic country rock (Fig. 1), but here we consider compaction of porosity within the fault zone. For simplicity, we treat the deformation as linear viscous, although similar behaviour holds for more complex rheologies. It is assumed that strain within the fault zone is a combination of shear strain along the plane of the fault zone and uniaxial strain normal to the plane of the fault zone associated with the reduction of porosity. The former, more familiar process is driven by the shear traction τ on the edges of the fault plane. The shear

creep rate is

$$\eta \frac{\partial V_{2y}}{\partial x} = \tau \quad (2)$$

where η is shear viscosity, V_2 is the velocity of the rock, x is the coordinate perpendicular to the fault zone, and y is parallel to fault slip. The latter process is driven by the difference between the normal traction T_n on the edges of the fault zone and the fluid pressure within the fault zone. The compaction creep rate is

$$K \operatorname{div} \mathbf{V}_2 = (P_f - P_s) = K \frac{\partial f_1}{\partial t} \quad (3)$$

where K is bulk viscosity, P_s is the mean pressure in the rock obtained from P_f and the boundary condition between the fault zone and the country rock involving T_n , f_1 is porosity, and t is time.

For ductile mechanisms, including pressure solution, grains and grain-grain contacts within the fault zone do not distinguish between the two sources of stress. Both the bulk and shear viscosity thus depend on the same intrinsic (grain) viscosity η_i which, although constant in our calculations, may depend non-linearly on stress, fluid pressure and porosity. Shear creep thus implies compaction creep. If the fault does not leak enough fluid to its surroundings on the timescale of an earthquake cycle, fluid pressure will increase and frictional faulting will occur at low shear tractions. The physics is similar to that of the segregation of melt from igneous mush and compaction of sediments, problems which have been treated in detail^{44–46}.

Changes in fluid pressure depend on the relative volume change of the fluid and hence the relative compaction rate of pore space. We envisage the porosity to consist of both equidimensional and high-aspect-ratio (tabular) voids, which we will refer to as 'pores' and 'cracks', respectively (Box 1). The compaction rate of pores is proportional to porosity, implying that changes in fluid pressure associated with them are nearly independent of porosity. The compaction rate of cracks is nearly independent of their porosity such that they are rapidly closed by creep. Overall, considerable fluid pressure changes are expected for any reasonable porosity within a creeping and mostly sealed fault zone.

Elastic strain. The earthquake cycle is simply represented in one dimension. The shear stress τ and the elastic displacement D stored in the country rock are the same everywhere on the fault and related by

$$\tau = D\mu/L \quad (4)$$

where μ is the rigidity of the country rock and L is a length scale. In our computations the elastic displacement is incremented at the start of each time step and stresses calculated. Shear strain, which reduces shear stress (equation (2)), and compaction, which increases fluid pressure, are computed (equation (3)) in parallel. Water pressure as a function of volume is obtained from ref. 47. Porous flow into or out of the fault zone changes the pressure in the fault. The stress state in the fault zone is then checked for frictional failure (equation (1)). The stress drop during earthquakes is assumed to reduce the shear traction by a fraction of that present before failure; elastic displacement, shear stress, porosity and fluid pressure are adjusted in parallel.

Porosity produced by earthquakes. Compaction continually removes porosity. A process to create porosity in the fault zone is thus needed so that porosity is not removed to the point that the final isolated pore space is at the confining pressure P_s . Frictional faulting during earthquakes should produce both cracks and pores in low-porosity fault zones within hard rocks¹⁸. Intuitively, broken pieces do not fit back together. Experimentally, granular materials dilate at high shear strain rates but compact at low shear strain rates^{26,27}.

BOX 1 Mathematical formalism

Closure of cracks and pores

Tabular cracks and equidimensional pores behave differently during creep, because cracks close faster than pores. Ideally one should keep track of the evolution and aspect ratio of each pore and crack during creep and compaction, but this would require consideration of an anisotropic material and would obscure our simple mechanism. We therefore consider two classes of pores: tabular cracks with an aspect ratio α_c that is the inverse of the crack porosity f_c , and spherical or cylindrical pores with an aspect ratio of 1. The shear viscosity of the medium is thus (from equation (A4) of ref. 41)

$$\eta = \frac{9\eta_i}{9 + 4\alpha_c f_c + 4f_p} \quad (\text{B1})$$

where η_i is the intrinsic viscosity of the grains and f_p is the pore porosity. The bulk viscosity is (from equation (A3) of ref. 41)

$$K = \frac{3\eta_i}{2\alpha_c f_c + 2f_p} \quad (\text{B2})$$

The compaction rate is inversely proportional to K . Thus it is nearly independent of porosity as long as $\alpha_c f_c$ is roughly 1. This creates a model singularity when all the cracks close at the same time. Physically, irregularities on a crack surface are expected to touch when a crack is nearly closed and preclude an indefinite increase in aspect ratio. Also, any cracks with very high aspect ratio quickly become arrays of fluid inclusions²³. In the calculations, the crack aspect ratio was restricted to an upper value $\alpha_{\max} = 120$ to allow the final cracks to close in a numerically orderly manner.

Increase of porosity by earthquakes

We use a continuous relationship between faulting energy and low crack porosities by combining equations (5) and (6)

$$\Delta f_c = -(f_c - \phi_c) \left[1 - \exp\left(-\frac{\beta_c G_F}{T_n W(\phi_c - f_c)}\right) \right] \quad (\text{B4})$$

where β_c is the small fraction of the faulting energy that goes into creating cracks. The expression saturates to ϕ_c , the maximum crack porosity. An analogous expression was used for pores.

Porosity production by faulting is limited by high confining pressures. Following ref. 26 (equations (5), (6)), porosity production is considered proportional to the energy per area of the fault plane, G_F , from elastic strain release, which is

$$G_F = \frac{1}{2} [\tau_{\text{old}} D_{\text{old}} - \tau_{\text{new}} D_{\text{new}}] \quad (\text{5})$$

in the one-dimensional example. Porosity created on the time-scale of an earthquake is considered to act as voids because there is too little time for water to flow into the pores, or conversely, voids do not preclude slip by instantaneously reducing the fluid pressure everywhere in the fault zone. The complex dynamical relationship between frictional slip, porosity and fluid pressure is beyond the scope of this paper. The energy per area expended in creating porosity is thus roughly

$$G_P \approx T_n W \Delta f_i \quad (\text{6})$$

where W is the width of the fault zone and Δf_i the change in

porosity. There is also an upper limit to the porosity that can be created by faulting in a fractured material; for example, faulting reduces porosity in highly porous sandstones^{48,49}. That is, the porosity of the fault zone saturates to a limiting value for large amounts of coseismic slip. A convenient linear formula for porosity generation is given in equation B4 (Box 1).

Porous flow. In a dynamic steady state with repeated characteristic earthquakes, porosity destroyed by compaction and porosity created by fracture may balance. In this case, fluid flow between the fault and country rock adjusts so that there is no net flow to or from the fault zone over an earthquake cycle. Following Rice¹³, we envisage that the permeability in the fault zone is much greater than the lateral permeability into the country rock, and that the country rock is more permeable than the seals that bound the fault zone. Thus we consider the fault zone to be a finite, uniformly pressurized reservoir separated by a thin flow barrier from the country rock, an infinite reservoir at hydrostatic pressure. In real fault zones, seal permeability probably varies during the earthquake cycle^{35,38}. For simplicity, however, we linearly represent the seal by

$$\frac{\partial P_f}{\partial t} = -\frac{P_f - P_h}{t_h} \quad (\text{7})$$

where P_h is the hydrostatic pressure in the country rocks. The leak time t_h is not important provided that it is long compared with the earthquake cycle and short compared with the geological time for which a fault is active. For simplicity, it is held constant.

Model results

A one-dimensional calculation is sufficient to illustrate the interaction of creep and compaction. We ran our numerical models until steady state was approached and then chose a few representative earthquake cycles for presentation. We use parameters grossly applicable to the San Andreas system, but do not attempt to fine-tune the models to a particular locality (Table 1). The effective length L is scaled to give reasonable slip in earthquakes. The basic features of the results are not sensitive to the poorly constrained porosity parameters.

We can study the effects of two less constrained parameters, the width W and intrinsic viscosity η_i of the fault zone, by modelling. Three models (1–3) can be used to illustrate the differences between faults where slip is mainly coseismic and those where slip occurs mostly by creep (Fig. 2, and Tables 2 and 3). For these models fault width is increased by factors of 10 over a range appropriate to exhumed portions of the San Andreas system³⁵ and intrinsic viscosity is decreased by factors of 10 while other physical parameters are held constant.

Mostly seismic faults. Minor amounts of ductile creep are sufficient to increase fluid pressure greatly and thus weaken the fault for earthquakes. The fault zone in model 1 (Fig. 2, left) is narrow and has a high viscosity. Thus creep accommodates only a very small fraction of the total slip. The small rate of shear creep strain of $3 \times 10^{-12} \text{ s}^{-1}$, however, is enough to weaken the fault through the sequence of events described above:

TABLE 3 Results

Model	Creep (%)	Cycle time (yr)	Shear traction, τ (MPa)*	Fluid pressure, P_f (MPa)†	Total porosity†, f_1	Pore porosity†, f_p	Seismic slip (m)
1	0.2	130	93.7	267–54	0.074–0.087	0.069–0.070	6.48
2	12.8	124	75.6	292–B	0.0083–0.0115	0.0083–0.0099	5.24
3	81.9	34	4.5	394–B	$(1.7\text{--}2.3) \times 10^{-6}$	$(1.7\text{--}2.3) \times 10^{-6}$	0.32

* Before earthquake.

† Before and after earthquake.

B, fluid pressure at boiling point.

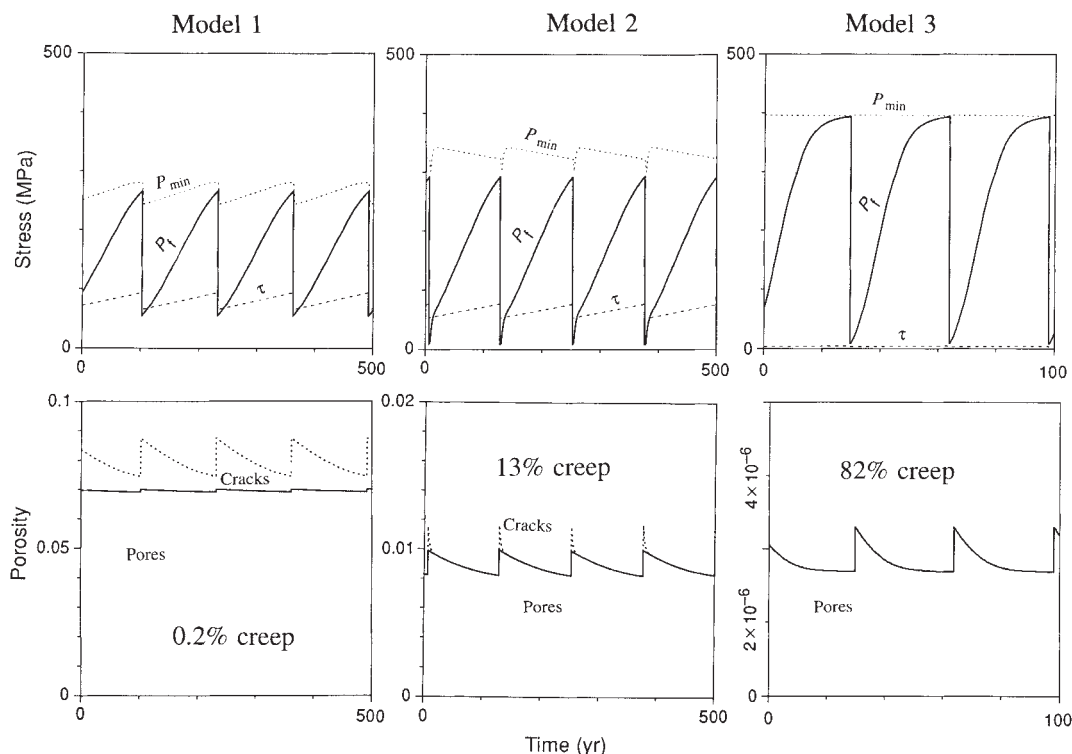


FIG. 2 Results of models 1–3 for a few earthquake cycles after steady state has been approached. Note the scale changes for the time and porosity axes. Fluid pressure P_f increases gradually between earthquakes while the shear traction on the fault τ increases. Frictional earthquake failure increases porosity and decreases shear traction and fluid pressure. The fluid pressure

fluctuates around the hydrostatic pressure, 140 MPa. Hydrofracture conditions do not occur because the least compressive stress in the fault zone $P_{\min}$ remains above P_f . Cracks compact much more easily than pores. The irregular behaviour of $P_{\min}$ in model 2 is associated with the closure of cracks early in each cycle.

high-aspect-ratio cracks compact between earthquakes, increasing the pore pressure. Fracture associated with earthquakes restores porosity and decreases fluid pressure. Following each earthquake fluid initially flows from the country rock into the underpressured fault zone despite ongoing compaction. Later, when P_f exceeds P_h , the direction of flow reverses.

A few dimensional parameters are helpful for interpreting the model. First, the maximum shear traction τ_{hydro} for driving creep within the fault zone is defined by the failure limit for hydrostatic pressure in equation (1). The creep velocity is thus less than

$$V_C = \tau_{\text{hydro}} W / \eta \quad (8)$$

which is 0.2 mm yr^{-1} for model 1. Second, the maximum length of an earthquake cycle in the absence of creep is the time for the stress to build up from the post-earthquake stress to the failure stress τ_{hydro}

$$t_f = L \tau_{\text{hydro}} \gamma / (\mu V_f) \quad (9)$$

where γ is fractional stress drop and V_f is the long-term velocity of the fault. This time is 250 yr for the parameters assumed in model 1. Third, the earthquake recurrence interval is shortened by the build-up of pore pressure. The time for pores to compact a significant fraction of their volume, dimensionally

$$t_p \approx \frac{\eta_i}{T_m - P_h} \quad (10)$$

is 3,850 yr, which is compatible with the slight pore closure that occurs in model 1. The analogous time for cracks

$$t_c \approx \frac{\eta_i f_c}{T_n - P_h} \quad (11)$$

where f_c is crack porosity, is 80 yr, compatible with the significant reduction in cracks and the shortened cycle time of 130 yr. The slow rate of pore closure limits the rate at which crack closure can increase fluid pressure.

The fault zone in model 1 is narrow; therefore the earthquake provides more than enough energy to regenerate porosity (see Box 1). In contrast, the fault zone in model 2 is wide enough that energy considerations limit porosity production. Cracks close early in the earthquake cycle and pores close slowly thereafter (Fig. 2, centre). Much of the fluid pressure increase in model 2 is thus controlled by pores.

Mostly creeping faults. Some faults, notably the Central California segment of the San Andreas, deform extensively by creep. The behaviour of a creeping fault is illustrated by model 3 which creeps with a strain rate of $1.3 \times 10^{-11} \text{ s}^{-1}$ (Fig. 2, right). Pores are included, but not high-aspect-ratio cracks; trial models indicated that cracks closed so rapidly that they had little effect on the earthquake cycle, but are numerically awkward because very short time steps are needed to resolve them. Pores in model 3 compact rapidly when the fluid pressure is low and then more slowly as the fluid pressure approaches T_n . In the model shown, a small earthquake occurs with a recurrence time of just 34 years. Failure of an actual fault might be triggered (or suppressed) by movement of adjoining segments of the fault or nearby branches of the fault system because the fault is near the failure condition throughout much of the cycle.

The viscosity is low enough in model 3 that fluid pressure increases to nearly lithostatic before much shear traction builds up. In this situation the cycle time is proportional to t_p , and the fraction of slip occurring by creep depends mainly on the width of the fault zone rather than its viscosity. Dimensionally, the shear traction at failure is

$$\tau_p \approx \frac{\mu V_f t_p}{L \gamma} \approx \frac{\mu V_f \eta_i}{L \gamma (T_n - P_h)} \quad (12)$$

The creep velocity at this shear traction

$$V_p \approx \frac{\tau_p W}{\eta} \approx \frac{\mu V_f W}{L \gamma (T_n - P_h)} \frac{\eta_i}{\eta} \quad (13)$$

is independent of viscosity because the shear viscosity η in equation (B1) is proportional to η_i . Thus narrow fault zones with low viscosity are expected to fail frequently at low shear tractions, but creep is expected to account for only a small fraction of the total slip.

Two difficulties arise in applying model 3 to the Earth. First, the computed porosity of 2×10^{-6} may be too low to enhance frictional failure. It might be better thought of as 2% porosity along a 1-cm-wide zone of the last few fault breaks. The porosity in the rest of the creeping zone would be unconnected fluid inclusions essentially at the confining pressure. Second, small amounts of connected but uncompactable porosity may delay the increase of pore pressure as do the nearly uncompactable pores in model 1 where cracks closed. For example, some pores may be surrounded by stronger minerals, or fluid-filled tubes may exist with serpentine. In this case, the slow rate of fluid pressure increase may allow infrequent, large earthquakes.

Implications

We have demonstrated that small amounts of ductile creep allow porous fault zones to compact. In a partially sealed fault zone, this increases fluid pressure so that frictional earthquake failure occurs at low shear tractions. The mechanism is likely to have broad application because it is probably self-adjusting in two ways. First, the porosity adjusts so that the amount generated during earthquakes is balanced by the amount lost to compaction. This allows the average fluid pressure within the fault zone to be in long-term equilibrium with hydrostatic pressure in the country rock. Second, frictional faulting creates extremely fine-grained material which is prone to ductile creep.

Low-permeability seals between the fault zone and the country rock are an integral part of the model. The seals need not be perfect, however, and water need not be trapped for geological times within a fault zone. In fact, a slightly leaky seal allows the average pressure during an earthquake cycle to be roughly hydrostatic. Over time, large volumes of meteoric water thus may enter and circulate through the fault zone, as is indicated by isotopic evidence from the San Andreas system³⁷. A deep crustal or mantle source of high-pressure water is not necessary.

The reproducible behaviour of earthquakes in the one-dimensional model indicates that earthquake prediction might be improved by considering both factors involved: shear traction and the increase of fluid pressure by ductile compaction. Shear traction is spatially continuous and its build-up during the earthquake cycle is monitored by available surface and geophysical techniques, whereas fluid pressure within fault zones is much more difficult to monitor. In practice, the rate of ductile creep is influenced by many factors including grain size, the geometry of porosity, temperature, rock composition and fluid composition. As noted above, minerals precipitated in fault zones may have important effects. At present, the physics of ductile creep and alteration of fault zones are not well enough understood to predict behaviour at depth from rock types exposed at the surface. It may be possible to calibrate compaction behaviour where a long enough record exists. Finally, the small volumes of fluid and rock in the 1-m-wide cores of major fault zones might be modified to be more tolerable neighbours by injecting fluids that enhance creep or fluids saturated with solids that creep easily, such as anhydrite or halite. □

Received 13 July; accepted 23 September 1992.

- Zoback, M. D. *Phil. Trans. R. Soc. A* **337**, 141–150 (1991).
- Hickman, S. H. *Rev. Geophys.* **29** (suppl.), 759–775 (1991).
- Lachenbruch, A. H. & Sass, J. H. *J. geophys. Res.* **85**, 6185–6222 (1980).
- Matthews, M. V. & Segall, P. *J. geophys. Res.* (in the press).
- Lisowski, M., Prescott, W. H., Savage, J. C. & Johnston, M. J. *Geophys. Res. Lett.* **17**, 1437–1440 (1990).
- Byerlee, J. D. *Pure appl. Geophys.* **116**, 615–626 (1978).
- Morrow, C., Radney, B. & Byerlee, J. D. in *Fault Mechanics and Transport Properties in Rocks (the Brace Volume)* (eds Evans, B. & Wong, T. F.) 69–88 (Academic, London, 1992).
- Cox, S. F. & Etheridge, M. A. *J. struct. Geol.* **11**, 147–162 (1989).
- Wannamaker, P. E. et al. *J. geophys. Res.* **94**, 14127–14144 (1989).
- Byerlee, J. D. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
- McCaig, A. M. *Geology* **16**, 867–870 (1988).
- Dick, H. J. B. et al. *Proc. ODP* **118**, 439–538 (1991).
- Rice, J. R. in *Fault Mechanics and Transport Properties in Rocks (the Brace Volume)* (eds Evans, B. & Wong, T. F.) 475–503 (Academic, London, 1992).
- Melosh, H. J. *J. geophys. Res.* **84**, 7513–7520 (1979).
- Lachenbruch, A. H. *J. geophys. Res.* **85**, 6097–6112 (1980).
- Raleigh, C. B. in *Proc. Conf. II, Experimental Studies of Rock Friction with Application to Earthquake Prediction*, 291–304 (U.S. Geological Survey, Menlo Park, California, 1977).
- McKenzie, D. P. & Brune, J. N. *Geophys. J. R. astr. Soc.* **29**, 65–78 (1972).
- Mead, W. J. *J. Geol.* **33**, 685–698 (1925).
- Rutter, E. H. *J. geol. Soc. Lond.* **140**, 725–740 (1983).
- Tada, R., Maliva, R. & Siever, R. *Geochim. cosmochim. Acta* **51**, 2295–2301 (1987).
- Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Geophys. Res. Lett.* **18**, 609–612 (1991).
- Lockner, D. A. thesis, Massachusetts Inst. Technol. (1990).
- Smith, D. L. & Evans, B. *J. geophys. Res.* **89**, 4125–4135 (1984).
- Morrow, C. A., Moore, D. E. & Byerlee, J. D. *Mater. Res. Soc. Proc.* **44**, 467–473 (1985).

- Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Nature* **358**, 574–576 (1992).
- Marone, C., Raleigh, C. B. & Scholz, C. H. *J. geophys. Res.* **95**, 7007–7025 (1990).
- Morrow, C. A. & Byerlee, J. D. *J. struct. Geol.* **11**, 815–825 (1989).
- Powley, D. E. *Earth Sci. Rev.* **29**, 215–226 (1990).
- Walder, J. & Nur, A. *J. geophys. Res.* **89**, 11539–11548 (1984).
- Hobbs, B. E., Ord, A. & Teyssier, C. *Pure appl. Geophys.* **124**, 309–336 (1986).
- Sibson, R. H., Robert, F. & Poulsen, K. H. *Geology* **16**, 551–555 (1988).
- Power, W. L. & Tullis, T. E. *J. struct. Geol.* **11**, 879–893 (1989).
- Parry, W. T. & Bruhn, R. L. *J. geophys. Res.* **91**, 730–744 (1986).
- Janecke, S. U. & Evans, J. P. *Geology* **16**, 1064–1067 (1988).
- Chester, F. M., Evans, J. P. & Biegel, R. L. *J. geophys. Res.* (in the press).
- Sibson, R. H. *Pure appl. Geophys.* **124**, 159–175 (1986).
- O'Neil, J. R. *Pure appl. Geophys.* **122**, 440–446 (1985).
- Byerlee, J. *Geology* (submitted).
- Byerlee, J. *Tectonophysics* (in the press).
- Angevine, C. L., Turcotte, D. L. & Furnish, M. D. *Tectonics* **1**, 151–160 (1982).
- Sleep, N. H. *J. geophys. Res.* **93**, 10255–10272 (1988).
- Jaeger, J. C. & Cook, N. G. W. *Fundamentals of Rock Mechanics* (Chapman and Hall, London, 1971).
- Sibson, R. H. *Bull. seismol. Soc. Am.* **80**, 1580–1604 (1990).
- McKenzie, D. P. *J. Petrol.* **25**, 713–765 (1984).
- McKenzie, D. P. *J. geol. Soc. Lond.* **144**, 299–307 (1987).
- Stevenson, D. J. & Scott, D. R. A. *Rev. Fluid Mech.* **23**, 305–339 (1991).
- Haar, L., Gallagher, J. S. & Kell, G. S. *NBS/NRC Steam Tables* (Hemisphere, Washington, 1971).
- Aydin, A. *Pure appl. Geophys.* **116**, 913–930 (1978).
- Aydin, A. & Johnston, A. M. *Pure appl. Geophys.* **116**, 931–942 (1978).

ACKNOWLEDGEMENTS. We thank M. L. Zoback, J. Savage, D. Lockner, J. Byerlee, P. Segall, G. Beroza, F. Chester, A. Aydin and G. Thompson. This work is funded by the U.S. Geological Survey and the U.S. NSF.

Targeted disruption of the mouse transforming growth factor- β 1 gene results in multifocal inflammatory disease

Marcia M. Shull*, Ilona Ormsby*, Ann B. Kier†, Sharon Pawlowski*, Ronald J. Diebold*, Moying Yin*, Ruth Allen*, Charles Sidman*, Gabriele Proetzel*, Dawn Calvin†, Nikki Annunziata† & Thomas Doetschman*

* Department of Molecular Genetics, Biochemistry, and Microbiology and † Division of Comparative Pathology, Department of Pathology and Laboratory Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524, USA

Transforming growth factor- β 1 (TGF- β 1) is a multifunctional growth factor that has profound regulatory effects on many developmental and physiological processes. Disruption of the TGF- β 1 gene by homologous recombination in murine embryonic stem cells enables mice to be generated that carry the disrupted allele. Animals homozygous for the mutated TGF- β 1 allele show no gross developmental abnormalities, but about 20 days after birth they succumb to a wasting syndrome accompanied by a multifocal, mixed inflammatory cell response and tissue necrosis, leading to organ failure and death. TGF- β 1-deficient mice may be valuable models for human immune and inflammatory disorders, including autoimmune diseases, transplant rejection and graft versus host reactions.

TRANSFORMING growth factor- β 1 is the prototypic member of a family of structurally related polypeptides referred to as the transforming growth factor superfamily. Members of this family exhibit a variety of proliferative, inductive and regulatory properties, and include the TGF- β s, activins, inhibins, bone morphogenetic proteins, mullerian inhibiting substance, *Drosophila* decapentaplegic gene complex, and *Xenopus* *Vg-1* gene¹⁻³. TGF- β 1 is synthesized as a precursor polypeptide containing a hydrophobic signal sequence, pro-region and mature peptide. The biologically active growth factor is a disulphide-linked homodimer, with each monomer representing the carboxy-terminal 112 amino acids of the precursor which are cleaved from the amino-terminal glycopeptide at a tetrabasic cleavage site¹⁻³.

Three TGF- β isoforms, termed TGF- β 1, TGF- β 2 and TGF- β 3, have been identified in mammals. These isoforms exhibit both overlapping and distinct spatial and temporal patterns of expression throughout development. Pronounced embryonic expression of the TGF- β s in areas undergoing morphogenetic events, particularly those involving epithelial-mesenchymal interactions, suggests that these molecules play a critical role during embryonic development^{4,5}. The large number of cells and tissues that express embryonic TGF- β 1 messenger RNA or protein are described in refs 6-11. In the adult, TGF- β 1 immunoreactivity is detected in many cell types¹².

TGF- β 1 elicits diverse cellular responses depending on cell type, state of differentiation and culture conditions¹⁻³. Thus, TGF- β 1 can have either stimulatory or inhibitory effects on the same cell, depending on the cellular environment¹⁻³. Biological actions of TGF- β 1 include regulation of cell proliferation, control of extracellular matrix protein production and degradation, and modulation of cellular differentiation¹⁻³. *In vitro* studies with TGF- β 1 have demonstrated inhibition of adipogenesis and myogenesis, inhibition of haematopoietic progenitor cell proliferation, promotion of angiogenesis, stimulation of chondrogenesis and osteogenesis, induction of bronchial epithelial cell differentiation, inhibition of intestinal epithelial cell growth, inhibition of keratinocyte proliferation, and modulation of numerous immune and inflammatory responses¹⁻³. The relevance of these multiple activities to normal development and functioning of the whole animal are unknown.

To investigate the role of TGF- β 1 *in vivo*, we disrupted one copy of the TGF- β 1 gene in murine embryonic stem (ES) cells by homologous recombination and generated chimaeric mice carrying the disrupted TGF- β 1 gene in the germ line. About 20 days after birth, animals homozygous for the mutated TGF- β 1 allele exhibit an acute wasting syndrome followed by death. No gross developmental abnormalities are detected. The most prominent lesions observed are tissue necrosis in specific organs and multifocal, mixed inflammatory cell infiltration into numerous organs, particularly heart and stomach.

TGF- β 1 gene targeting

To ensure that strain differences between the ES cell and targeting DNA did not affect targeting efficiency, the murine TGF- β 1 gene was isolated from a genomic library constructed with DNA isolated from a 129/SvJ mouse ($A^w/A^w c^{ch}/c p/p$), which is coisogenic to the 129/Sv strain ($A^w/A^w +^c/+^c +^p/+^p$) from which the D3 ES cell line¹³ used here was derived. The targeting strategy was designed to effect disruption of the mature peptide, while leaving the precursor pro-region intact. The targeting vector consisted of 4.0 kb of genomic TGF- β 1 sequence containing a *neo*^r gene inserted into exon 6 (Fig. 1). The purified targeting fragment was introduced into D3 cells by electroporation, and the cells were grown in the presence of G418, which selects for neomycin resistance, for 12-14 days. ES cells containing a disrupted TGF- β 1 gene were identified by polymerase chain reaction (PCR) and Southern blot hybridization of the PCR products using a TGF- β 1 oligonucleotide probe spanning the 3' insertion site. Four individual PCR-positive clones were obtained from a total of 700 G418-resistant colonies. Genomic Southern blot analysis using a TGF- β 1 genomic fragment as probe (probe B; Fig. 1) confirmed that three of these clones contained the desired homologous recombination event, yielding an overall targeting efficiency of 1 in 2×10^6 electroporated cells. As shown in Fig. 1, cell lines 1-1, 9-1, and 14-4 were heterozygous, containing one copy of the disrupted TGF- β 1 allele and one copy of the normal allele. Analysis of these cell lines using five additional enzymes and using both a *neo* probe (probe A; Fig. 1) and a probe external to the targeting fragment (probe C; Fig. 1) also confirmed that accurate targeting had occurred (data not shown).

Chimaeric mice were generated by injecting ES cells from lines 9-1 and 14-4 into C57BL/6J blastocysts and implanting the blastocysts into the uteri of pseudopregnant F₁ (C3H/HeN × C57BL/6J) recipients. Ten fertile coat colour chimaeric males were obtained, five of which exhibited >50% agouti coat colour. Mating of these five animals to strain CF1 females allowed identification of three animals that transmitted the ES cell-derived agouti coat colour to their offspring. Two of these males have produced a total of 18 litters and have transmitted the modified TGF- β 1 allele to 50% of their agouti offspring, as expected. Heterozygous mice carrying one wild-type and one disrupted TGF- β 1 allele are phenotypically normal and fertile. Heterozygous animals derived from both cell lines have been intercrossed to generate animals homozygous for the altered TGF- β 1 allele. Genotypes of the progeny were determined by Southern blot analysis of tail DNA biopsies taken at roughly three weeks of age. A representative example of such a litter is shown in Fig. 2. Of 117 offspring from 14 heterozygote crosses, 34 were homozygous wild-type, 67 were heterozygous, and 16 were homozygous for the disrupted allele. A statistically significant ($P < 0.005$) deviation from the expected 1:2:1 ratio of homozygous wild-type, heterozygous, and homozygous mutant progeny was observed, because of an under-representation of the homozygous mutant class of animals. Here we report analyses of live homozygous mutant animals derived from targeted cell line 14-4. Homozygous mutant animals derived from cell

line 9-1 exhibited the same phenotype as those derived from line 14-4.

Gross phenotype of homozygous mutants

Mice homozygous for the disrupted TGF- β 1 allele were phenotypically indistinguishable from wild-type or heterozygous littermates until roughly three weeks after birth. At this time, all homozygotes exhibited severe wasting followed by death (or if they seemed so ill that survival until the next day was doubtful, they were killed). In some animals, wasting was accompanied by dishevelled appearance, hunched posture, conjunctivitis, and skin irritation. The age at which symptoms first appeared was variable, ranging from 17 to 34 days. Generally, progression of the illness was rapid, with initial onset of symptoms to death lasting only a few days. In one animal, number 181, the rapidity of the illness was dramatic, with no indication of illness before the day the animal died. Gross anatomical examination of animals homozygous for the mutant TGF- β 1 allele revealed no apparent developmental abnormalities. But in most mutant animals, the spleens were smaller and Peyer's patches less prevalent than those of normal littermates. This phenotype was not observed in any homozygous wild-type or heterozygous littermates, nor was it seen in any of 55 progeny resulting from interbreeding agouti animals which were homozygous wild-type at the TGF- β 1 locus, that is, they had received the ES cell-derived unmodified TGF- β 1 allele.

FIG. 1 Targeted disruption of the murine TGF- β 1 gene in ES cells. **a**, Targeting construct consisting of a TGF- β 1 4.0-kb *Sma*I genomic fragment containing exon 6 and part of exon 7. A *neo*^r gene lacking the polyadenylation signal (*Xho*I-*Sal*I fragment from pMC1 *neo*; ref. 52) was inserted into the *Bam*HI site in exon 6 of TGF- β 1, 102 nucleotides (34 amino acids) from the N terminus of the mature peptide. **b**, Restriction map of the wild-type TGF- β 1 genomic locus surrounding the targeting site. **c**, Predicted structure of the disrupted TGF- β 1 allele. Arrowheads represent positions of primers used for PCR analysis. The 5' primer (5'-GCTTT-ACGGTATCGCCGCTC-3') is located 68 nucleotides upstream of the stop codon in *neo*. The 3' primer (5'-TGCGA-CCCACGTAGTAGACG-3') is located 74 nucleotides downstream of the *Sma*I site in exon 7 of TGF- β 1 and is not contained in the targeting vector. The locations of probes used in Southern analysis are shown. Probe A is from the *neo*^r gene (*Xho*I-*Sal*I fragment from pMC1 *neo*). Probe B is a 0.6-kb TGF- β 1 genomic fragment (*Bam*HI-*Sma*I). Probe C is a 0.7-kb TGF- β 1 genomic fragment (*Eco*RI) not contained in the targeting vector. Restriction enzymes used in diagnostic digests are: A, *Apal*; Bs, *Bst*I; E, *Eco*RI; N, *Nco*I; P, *Pst*I; Pv, *Pvu*II; Sp, *Sph*I; S, *Stu*I; and X, *Xba*I. **d**, Southern blot analysis of parental D3 embryonic stem (ESD3) cells and three PCR-positive clones. Genomic DNAs (15 μ g) were digested with the enzymes shown, electrophoresed through 0.8% agarose, transferred to nylon membranes, and hybridized with probe B. The observed patterns correspond to those expected from homologous recombination of the targeting vector into the TGF- β 1 locus. The difference in intensity of the bands in the digests of clone 1-1 is due to the fact that this clone contained a mixed population of targeted and untransfected cells.

METHODS. D3 embryonic stem (6×10^6 cells) were electroporated with 10 μ g (6 nM) of purified targeting fragment using an IBI Gene Zapper at 800 V cm^{-1} and 200 μ F. Cells were plated on mitomycin-treated *neo*^r mouse embryonic fibroblast feeder layers. After 24 h with no selection, 400 μ g ml^{-1} G418 was added. After further 20 h, the medium was changed to 200 μ g ml^{-1} G418 and cells were grown in this medium for 12–14 days. G418-resistant colonies (initially in pools and then individual colonies from PCR-positive pools) were analysed for homologous recombination by PCR. Targeting was confirmed by genomic Southern blot analysis.

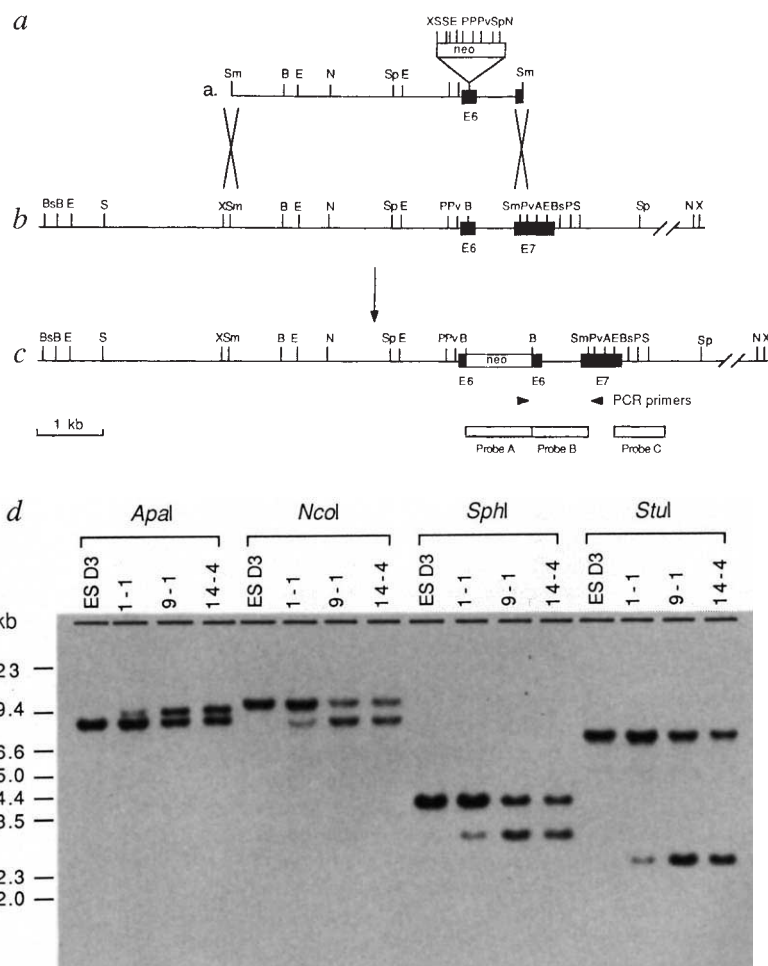


TABLE 1 Histological analysis of TGF- β 1-deficient mice

Mouse	Control*†	126	137	147	165	166	181	200	208	254	258	260	366	374	379	397
Genotype	+/-, +/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
Sex	F, M	F	F	F	M	F	M	F	F	M	F	M	M	F	F	M
Age at death (days)	21-37†	29†	31†	21	25†	22	27	19†	20†	21†	34†	21†	23	23†	23†	21†
Weight at death (g)	11-22	ND	ND	ND	8.0	7.0	12.0	6.0	7.0	8.0	9.8	11.8	9.1	7.5	7.9	8.5
Heart (inflammation)																
Pericardium	-	+/-	-	-	-	+	++	-	-	++	++++	+	+	+	-	+
Myocardium																
Atria	-	+/-	+	+	+	+	+++	+/-	+	++	++++	+	+++	++	++	++
Ventricles	-	-	-	-	+	-	+++	-	-	++	++	+	+	+	-	+
Endocardium	-	+/-	-	+	+	-	+++	-	+/-	+	+	+	+/-	++	+/-	+
Lung																
Perivascularitis	-	-	-	-	-	-	+	+	++/-	++	+/-	++	+	+	++/-	+/-
Perivascular oedema	-	-	-	-	-	-	+/-	-	-	-	-	-	-	-	-	-
Stomach																
Inflammation	-	+++	++++	++++	++++	++	++	+++	-	++	++	++	pmd	+/-	+++	++++
Ulcer	-	+++	++++	++++	++++	-	pmd	+++	-	-	-	-	pmd	-	+++	+
Acanthosis	-	++	++	++	++	++	+	++	-	++	++++	+++	pmd	+/-	+/-	++++
Hyperplasia	-	-	-	-	++	++	-	++	-	+++	++	+++	pmd	+/-	+	+++
Liver																
Inflammation	-	+++	++	+	+++	+	++	+++	+	++/-	+	++/-	+	+	++/-	++
Microgranulomas	-	++	++	-	-	+	-	++	-	-	-	-	-	-	-	-
Pancreas (inflammation)	-	-	+	+/-	+++	-	-	-	++	++	ND	++	pmd	-	+++	+
Salivary gland (inflammation)	-	-	-	-	+/-	-	++	-	-	-	+	++	+	-	+/-	++/-
Spleen																
Size smaller	N	Y	Y	Y	N	Y	Y	ND	Y	N	Y	Y	pmd	Y	Y	N
Indistinct white pulp	N	Y	Y	Y	N	Y	Y	ND	Y	N	Y	Y	pmd	N	N	Y/N
Lymph nodes																
Mediastinal	-	-	lg	lg	-	-	-	-	-	-	-	-	lg	-	-	lg
Mandibular	-	-	-	lg	-	-	-	lg	-	-	-	lg	-	lg	lg	-
Mesenteric	-	lg	-	lg	-	-	-	lg	-	lg	-	-	-	-	lg	-
Peyer's patches‡	4-14	1	2	pmd	2	7	pmd	2	0	4	2	3	pmd	2	8	3
Brain (inflammation)	-	ND	ND	-	-	-	-	+/-	ND	-	-	+/-	-	-	-	-
Eye (inflammation)																
Conjunctiva	-	ND	ND	-	ND	+	ND	ND	ND	++	-	+/-	ND	-	ND	+
Lacrimal glands	-	ND	ND	-	ND	-	ND	ND	ND	-	-	-	ND	-	ND	++
Striated muscle	-	ND	ND	-	ND	-	ND	ND	ND	++	-	+	ND	-	ND	+
Striated muscle (inflammation)	-	-	ND	-	+	-	+	ND	-	+	+	++	ND	ND	+++	+++
Serosa (inflammation)	-	+/-	ND	ND	ND	+/-	ND	ND	++	+++	+	++	pmd	++	-	+/-
Haematocrit§	24-50	ND	ND	ND	ND	ND	ND	ND	ND	27	52	24	ND	46	30	35.5

Degree of effect is indicated by - (no difference from control) to ++++ (large effect). N, no; Y, yes; lg, large; ND, not done; pmd, post mortem degeneration.

* Summary of results from three homozygous wild-type (3 females, aged 23, 29, and 31 days) and three heterozygous normal (1 female, 34 days; 2 males, 21 days) littermates.

† Animals were killed.

‡ Number of Peyer's patches/total length of intestine.

§ Leukocyte haematology is given in Table 2.

These observations indicate that the mutant phenotype was due to homozygosity of the targeted TGF- β 1 gene, and not to homozygosity of other D3 cell genes.

Histopathology

Histological analyses were performed on 15 homozygous mutant animals and six normal wild-type or heterozygous littermate controls (Table 1). Whereas the normal littermate controls exhibited no remarkable histopathology, all homozygous mutant animals exhibited a marked but variable degree of mixed inflammatory cell infiltration and tissue necrosis in several organs. The main organs involved were heart, stomach, liver, lung, pancreas, salivary gland and striated muscle. In liver, pancreas and salivary gland, the inflammatory cell infiltration was primarily periductal. The stomach inflammation appeared to be primarily granulocytic (~60% neutrophils and some eosinophils), whereas heart, lung, pancreas, striated muscle, salivary gland and brain infiltrates appeared primarily (~80%) lymphocytic and plasmacytic. In liver, granulocytes and lymphocytes appeared to be present in roughly equal numbers. Representative examples of affected tissues are shown in Fig. 3.

The heart was affected in most homozygous mutant animals examined, although the degree of inflammation and necrosis was variable. In several cases, the presence of inflammatory cells was generalized and extensive, involving the pericardium, myocardium and endocardium of atria and ventricles (Fig. 3a and b). Striated muscle (masseter, leg and/or diaphragm) was examined in 11 mutant animals; eight of these exhibited some inflammation. In two animals (379 and 397), inflammation and necrosis of the diaphragm were severe enough to contribute to

breathing difficulty and possibly death. A generalized mild to moderate perivascularitis was observed in the lungs of 10 of 15 mutant animals (Fig. 3c).

The non-glandular portion of the stomach exhibited some degree of acanthosis and inflammatory cell infiltration into the submucosa in 13 of 14 homozygous mutant animals in which this organ could be examined. Severe focal ulceration (seven animals) and/or hyperplasia of the basal epithelial layer (nine animals) accompanied the inflammation. As shown in Fig. 3d, severe, extensive ulceration was observed in the junction between the glandular and nonglandular portion of the stomach and the submucosa was filled with mixed inflammatory cells, fibrin and focal haemorrhage.

Inflammatory cell infiltration in the portal triad area of the liver, with cells most concentrated near the bile ducts, was observed in all 15 mutant animals (Fig. 3e). Some animals also exhibited multifocal hepatic necrosis and microgranulomas. Mild to moderate inflammation of the serosa of internal organs (stomach, intestine, kidney, ovary and testis) was seen in several animals. In 8 of 13 animals, some degree of multifocal inflammation was observed in the pancreas, and in 7 of 15 animals, a slight to moderate multifocal inflammatory cell infiltration in the salivary gland was detected (Fig. 3f). None of the mutant animals exhibited any kidney inflammation other than that of the renal serosa.

Histological analysis of lymphoid organs revealed that, compared with normal littermates, 10 of 15 homozygous mutant animals had slightly enlarged lymph nodes and 10 of 13 animals had spleens that were smaller and with less distinct white pulp (Fig. 3g and h). Peyer's patches were usually fewer and

TABLE 2 Total numbers and differential distributions of blood leukocytes

Control mice*							Mutant mice						
		Total WBC	Neutrophils						Total WBC	Neutrophils			
Mouse†		(per mm ³)	Lymphocytes	Total	Immature	Monocytes	Mouse	(per mm ³)	Lymphocytes	Total	Immature	Monocytes	
362F	25(+/-)	2,400	1,896 (79)	432 (18)	0 (0)	24 (1)	166F	22(-/-)	1,600	800 (50)	368 (23)	96 (6)	384 (24)
451F	20(ND)	2,500	1,325 (53)	950 (38)	100 (4)	175 (7)	208F	20(-/-)	1,300	715 (55)	429 (33)	39 (3)	156 (12)
447M	20(ND)	2,800	1,568 (56)	952 (34)	84 (3)	196 (7)	254M	21(-/-)	2,000	1,180 (59)	400 (20)	100 (5)	320 (16)
455F	19(ND)	2,300	1,679 (73)	506 (22)	92 (4)	69 (3)	260F	21(-/-)	2,500	1,650 (66)	600 (24)	250 (10)	225 (9)
377F	23(+/-)	1,600	1,072 (67)	384 (24)	64 (4)	144 (9)	374F	23(-/-)	4,200	2,730 (65)	1,050 (25)	168 (4)	378 (9)
532F	20(ND)	2,100	1,428 (68)	630 (30)	63 (3)	42 (2)	379F	23(-/-)	5,500	3,245 (59)	1,925 (35)	275 (5)	275 (5)
392M	22(+/-)	2,500	1,825 (73)	500 (20)	175 (7)	150 (6)	397M	22(-/-)	2,500	1,225 (49)	1,050 (42)	750 (30)	225 (9)
425M	37(ND)	4,500	3,510 (78)	945 (21)	0 (0)	45 (1)	418M	37(-/-)	4,000	2,080 (52)	1,640 (41)	600 (15)	200 (5)
471F	29(+/-)	3,700	2,812 (76)	814 (22)	37 (1)	74 (2)	473F	29(-/-)	5,000	2,350 (47)	2,150 (43)	500 (10)	400 (8)
mean		2,711	1,902 (69)	679 (25)	68 (3)	102 (4)			3,178	1,775 (56)	1,068 (32)	309 (10)	285 (11)
±s.d.		±877	±777 (±9)	±237 (±7)	±54 (±2)	±64 (±3)			±1,531	±883 (±7)	±689 (±9)	±251 (±8)	±89 (±6)
<i>P</i> value‡							<i>P</i> >0.05		<i>P</i> >0.05	<i>P</i> >0.05	<i>P</i> <0.02	<i>P</i> <0.01	

Absolute number of cells per mm³ is given followed by percentage in parentheses. WBC, white blood cells.

* Control mice were sex, parent and, as closely as possible, age matched.

† The mouse number is followed by sex, age at death or sacrifice (days), and genotype (ND, not determined, phenotypically normal).

‡ t-test of paired comparisons between control and mutant animals.

possessed less distinct germinal centres than those of normal littermates. The thymus of mutant animals exhibited no significant lesions in the sections examined.

As one of the initial external symptoms in the homozygous mutant animals was conjunctivitis, the eyes of seven mutant animals were examined histologically, revealing conjunctivitis in four animals, inflammation of ocular striated muscle in three, and lacrimal gland inflammation in one. A slight focal inflammation in the brain choroid was observed in two animals.

Inflammation often accompanies infection and the inflammatory cell infiltration in some of the homozygous mutant animals resembled that seen in certain murine bacterial or viral diseases. However, the mutant phenotype was not observed in any wild-type or heterozygous littermates. Furthermore, mice were maintained in microisolator cages in a clean facility, with sentinel animals present for detection of antibodies against 14 murine bacterial and viral pathogens. Periodical comprehensive serology and histopathology revealed no indication of pathogens in the sentinels. Serum samples from two heterozygous animals producing litters containing homozygous mutant animals were assayed for the presence of antibodies against nine common murine bacterial and viral pathogens, including mouse hepatitis virus; no such antibodies were detected. Blood, liver, lung and fluid from intraperitoneal lavage from two mutants and one control animal were cultured in Columbia paediatric broth and analysed for the presence of gram-negative bacteria. No significant bacterial pathogens were detected. In addition, a C57BL/6 mouse doubly homozygous for the *scid* (severe com-

bined immune deficiency) mutation¹⁴ and a non-responder allele at the *Lps* (lipopolysaccharide response) locus¹⁵ has been housed with the germ-line chimaera for over a month and has remained healthy. These results suggest that the phenotype observed in homozygous mutant animals is not due to primary infection by common mouse pathogens, although the possibility of secondary bacterial or viral involvement in the mutant phenotype cannot be eliminated completely.

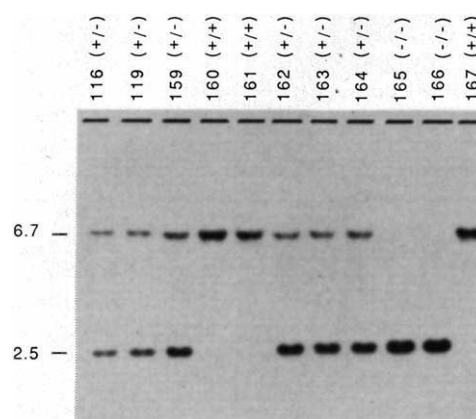
Haematology

Peripheral blood samples from 9 homozygous mutants and an equal number of age-, sex- and parent-matched homozygous wild-type or heterozygous controls were analysed for total numbers and differential distributions of leukocytes. In mutant animals, the average number of white blood cells was usually elevated compared with controls, the increase being due to greater absolute numbers of neutrophils and monocytes. Even in this small group of mice, a statistically significant increase in monocytes and immature forms of neutrophils was observed (Table 2), indicating that a responsive left shift was present in the mutant mice. Haematocrits, determined for four control and six mutant animals, revealed no significant differences between the mutants and controls (Table 1).

Cytokine analysis

The expression of cytokines important in immune and inflammatory responses was investigated by PCR analysis of mRNAs from spleen, liver and lung from one heterozygous normal and

FIG. 2 Genotype of offspring from interbreeding mice heterozygous for the targeted TGF- β 1 allele. Tail DNAs from parents (116 and 119) and offspring (159–167) were digested with *Stu*I, fractionated by electrophoresis through 0.8% agarose, transferred to nylon membranes, and hybridized with a TGF- β 1 probe (probe B; Fig. 1). The upper band (6.7 kb) represents the wild-type allele and the lower band (2.5 kb) represents the targeted allele. +/+, Homozygous wild-type; +/-, heterozygote; -/-, homozygous targeted.



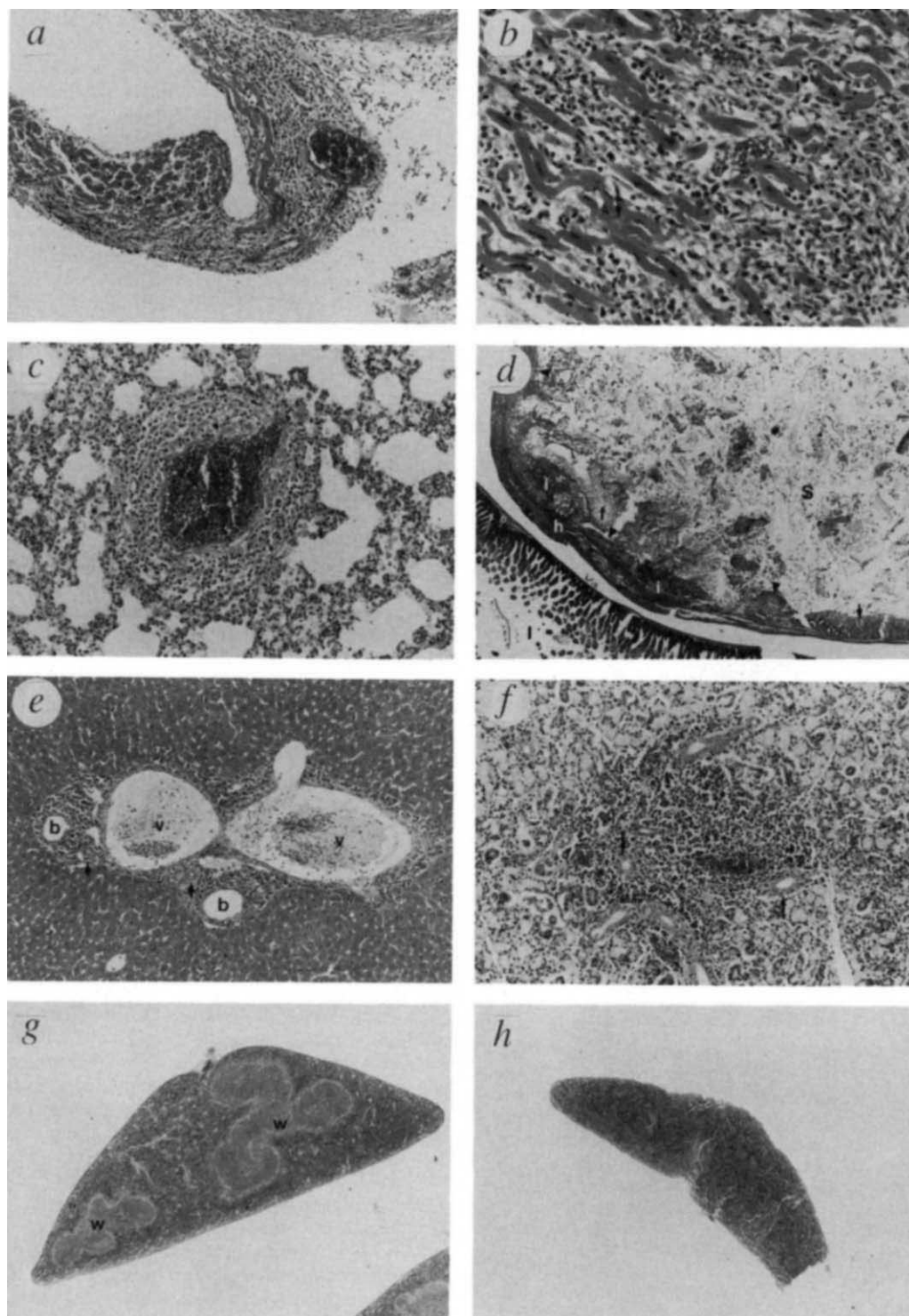


FIG. 3 Histological analysis of TGF- β 1 deficient mice. *a*, Heart, atrium, mouse 181, $\times 74$. Myocarditis and pericarditis (arrow). The myocardium is infiltrated with inflammatory cells in all areas of the heart. Single cell to focal myocyte necrosis is evident. *b*, Heart, higher magnification ($\times 148$) ventricle, mouse 181. Inflammatory cells consist mostly of lymphocytes, but neutrophils, macrophages, and plasma cells are also present. Individual myocyte necrosis is evident (arrow). *c*, Lung, mouse 181, $\times 93$. Perivascularitis. Inflammatory cells surround a vessel wall, with some oedema. The types and proportions of cells are similar to those seen in the heart. *d*, Stomach, gut roll, mouse 137, $\times 9$. Severe, generalized necrosis (ulceration) of the squamous epithelium in the junction between the glandular (arrow) and nonglandular (arrowheads) portion. The submucosa is filled with mixed acute inflammatory cells (i), particularly neutrophils, and fibrin (f), with focal hemorrhage (h).

Stomach lumen (S); intestine (I). *e*, Liver, mouse 126, $\times 46$. Typical pattern of acute inflammatory cell infiltration in the portal triads of the liver. Infiltrating cells are most concentrated near bile ducts and consist mostly of lymphocytes and neutrophils. Bile ducts (b); portal veins (v); arteries (arrows). *f*, Salivary gland, mutant 181, $\times 74$. Typical focal area of inflammatory cell infiltration seen particularly around ductal epithelium (arrows) of salivary glands and pancreas of some mutants. The inflammation is primarily lymphocytic. *g*, Spleen from normal littermate, 128, $\times 13$. White pulp areas (w) are large and clearly delineated. *h*, Spleen from mutant 147, $\times 13$. The white pulp areas are not detectable, and the entire spleen is smaller in cross-section. Tissues were prepared, sectioned, and stained with haematoxylin and eosin according to standard procedures.

two homozygous mutant animals (Fig. 4). As expected, no wild-type TGF- β 1 mRNA was detected in the mutant animals, although it was detected in all tissues of the control. Interferon- γ , tumour necrosis factor- α (TNF- α) and macrophage inflammatory protein-1 α (MIP-1 α), which are major mediators of inflammation, yielded strong amplification products in liver and lung of the mutant animals, but were not detected in these tissues in the control animal under the PCR conditions used. Interferon- γ , TNF- α and MIP-1 α were detected in the spleens of both the control and mutant mice, although in one mutant, the amounts of product amplified were greater than those of the control, and in the other, less. Interleukin-1 β , also an important mediator of inflammation, was present in all three tissues in both the control and mutant animals; in liver the amount of PCR product was slightly greater in the mutants than in the control. There were no significant differences between the control and mutant animals in the amounts of amplification products for TNF- β and interleukin-1 α (data not shown).

Discussion

Histological analysis of animals homozygous for the disrupted TGF- β 1 allele revealed a multifocal, mixed inflammatory cell

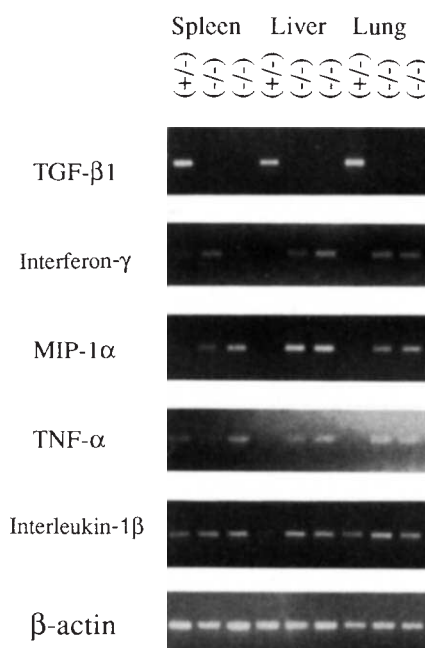


FIG. 4 PCR analysis of cytokine mRNAs in control and TGF- β 1-deficient mice. Total RNAs from spleen, liver, and lung from control mouse 249 (+/-) and homozygous mutant animals 254 and 260 (-/-) were reverse-transcribed and examined by PCR analysis using primers specific for cytokines TGF- β 1 (5'-GCGACTACTATGCTAAAGAGG-3' and 5'-GTTGTGTTGGTTGTAGAGGCA-3', 40 cycles), interferon- γ (5'-TGGCTGTTCTGGCTGTTACTG-3' and 5'-AATCAGCAGCGACTCCTTTTCC-3', 35 cycles), MIP-1 α (5'-ACTGCCCTGCTGTTCTTCT-3' and 5'-AGGCATTTCAGTCCAGGTCAGT-3', 40 cycles), TNF- α (5'-CCAGACCTCAGCTCAGAT-3' and 5'-AACACCCATTCCTTCACAG-3', 31 cycles), and interleukin-1 β (5'-TGACGGACCCAAAGATGAAG-3' and 5'-CTGCTTGAGGCTGATGTA-3', 35 cycles). Amplification conditions using a Perkin-Elmer 9600 were 95 °C, 30 s; 58 °C, 1 min; 72 °C, 2 min (extended 1 s per cycle). The final extension was allowed to continue for 10 min. Amplified products were size-fractionated by electrophoresis through agarose and visualized by ultraviolet illumination of the ethidium bromide-stained gel. With each set of samples (except TNF- α and β -actin), positive and negative PCR controls were run. The positive control consisted of total RNA from spleen cells cultured for 48 h with concanavalin A. The negative control contained all complementary DNA synthesis and PCR reagents but lacked RNA. Correct amplification products were observed in each positive control, whereas no products were amplified in the negative controls (data not shown). All samples gave comparable signals when amplified using primers specific for β -actin (5'-GTGGGCGCTCTAGGCACCAA-3' and 5'-CTCTTTGATGTCACGCACGATTC-3').

infiltration, often with necrosis. The infiltrates, which seemed to contain primarily lymphocytes and neutrophils, affected many organs and often seemed so severe as to cause organ compromise or failure. Haematological analysis revealed a responsive increase in monocytes and immature neutrophils in the mutant animals. Important cytokine mediators of inflammation, including interferon- γ , TNF- α , and MIP-1 α , were elevated in liver and lung of mutant animals relative to the control, consistent with the inflammation observed in these tissues. Thus, TGF- β 1 deficiency resulted in severe pathology leading to death associated with dysfunction of the immune and inflammatory systems.

TGF- β 1 is a potent regulator of the immune and inflammatory systems, generally functioning as an immunosuppressive agent. TGF- β 1 suppresses B-cell proliferation and immunoglobulin (IgG and IgM) secretion^{16,17}; inhibits thymocyte, T lymphocyte, and large granular lymphocyte proliferation¹⁸⁻²¹; suppresses natural and lymphokine-activated killing by large granular lymphocytes^{22,23}; and inhibits generation of cytotoxic T lymphocytes²⁴. Both the initiation and resolution of general inflammatory responses seem to involve TGF- β 1. TGF- β 1 stimulates monocyte migration and growth factor production²⁵. But after initiation of an inflammatory response, it also exhibits potent anti-inflammatory effects including inhibition of neutrophil and T-lymphocyte adhesion to endothelium²⁶, down-regulation of macrophages^{25,27} and antagonism of TNF- α function^{24,28}. Given these inhibitory effects, TGF- β 1 deficiency might perturb normal immune and inflammatory system homeostasis, resulting in unregulated activity that is fatal to the organism. For example, dysregulated production of interferon- γ and TNF- α by activated T cells and macrophages, respectively, could contribute to the mutant phenotype as both cytokines mediate weight loss and inflammation^{29,30}. The involvement of interferon- γ and TNF- α are consistent with the observed elevation of these cytokines in the mutant animals.

The inflammatory cell infiltration observed in some tissues of homozygous mutant animals resembles to a variable degree that seen in certain human autoimmune diseases such as myocarditis, polymyositis³¹, and Sjogren's syndrome³². Pathological autoreactivity has been attributed to several processes in which TGF- β 1 may play a regulatory role, including dysfunction of T cell suppression and aberrant expression of major histocompatibility complex (MHC) antigen by cells that normally do not present antigen³³⁻³⁵. TGF- β 1 administration to mice protects against collagen-induced arthritis and relapsing experimental allergic encephalomyelitis³⁶⁻³⁹. T-cell suppression of experimental autoimmune encephalomyelitis after oral tolerization to myelin basic protein seems to be mediated by TGF- β 1 (ref. 40), suggesting that the absence of TGF- β 1 may reduce suppression by T cells and lead to autoreactivity in the affected animal. The autoimmune state may involve interferon- γ -induced expression of MHC class II molecules in a variety of cells³⁴. TGF- β 1 downregulates interferon- γ -induced MHC class II antigen expression on both lymphoid and non-lymphoid cell types⁴¹ and is a potent inhibitor of interferon- γ production by peripheral blood mononuclear cells²⁸. Consequently, absence of TGF- β 1 could lead to presentation of self antigens by inappropriate cells, thereby eliciting an autoimmune response.

TGF- β 1 also seems to play an important role in regulating the mucosal immune system. Mucosal or secretory immunity involves induction of precursor IgA-secreting cells in bronchial-associated and gut-associated lymphatic tissue, including Peyer's patches; migration of these cells to mucosal sites (including gut, respiratory tract, mammary gland, lacrimal gland and genitourinary system); and terminal differentiation to IgA-secreting cells in these tissues⁴². TGF- β 1 specifically enhances IgA expression in stimulated Peyer's patch and splenic B cells^{43,44} and seems to be involved in lymphocyte homing to mucosal sites, enhancing expression of human HML-1 and murine β 7 α M290 integrins which are believed to mediate homing to the mucosa^{45,46}. On the other hand, TGF- β 1 inhibits

adhesiveness of Peyer's patch high endothelial venules for lymphocytes⁴⁷. TGF- β 1 deficiency could, therefore, result in dysfunction of the mucosal immune system, including abrogation of orally induced suppression, inhibition of IgA secretion and aberrant lymphocyte homing to mucosal sites, leading to adverse systemic consequences.

As the TGF- β s may play critical roles in embryogenesis¹⁻⁵, it is intriguing that animals homozygous for the disrupted TGF- β 1 gene exhibited no obvious developmental defects. It is proposed that functionally redundant genetic pathways may exist in critical developmental processes, involving structurally related genes or unrelated genes which have evolved overlapping functions⁴⁸. The existence of multiple TGF- β isoforms, often exhibiting similar biological activities and overlapping patterns of expression during development, raises the possibility that expression of these other isoforms compensates for the lack of TGF- β 1. But the under-representation of animals in the homozygous mutant class of progeny when heterozygous animals are interbred should be noted. Loss of half of the homozygous mutant animals seems to have occurred. The occurrence of this apparent embryonic lethality may depend on segregation of an additional gene(s) in the progeny and may reflect the genetic mix (129 and CF1) of the homozygous mutant animals.

An additional explanation for the failure to observe a developmental phenotype is that TGF- β 1 is, in fact, required for embryonic and postnatal development and viability, but that maternal rescue of the TGF- β 1-deficient animals occurs *in utero* by means of transplacental transfer of the growth factor and postnatally by transmission in the milk. TGF- β 1 was first isolated from

placenta⁴⁹, but whether transplacental transfer occurs has not been determined. TGF- β -like activity has been detected in human and bovine colostrum⁵⁰, and a TGF- β 2-like protein has been isolated from bovine milk⁵¹. Thus, the onset of symptoms at the time of weaning may reflect maternal rescue before weaning. If this is the case, then this system provides a unique opportunity to study, in the maturing animal, the effects of a growth factor which is required during embryogenesis but which also has distinct functions in the adult.

Studies are in progress to determine the precise nature and mechanism of the immune and inflammatory dysfunction resulting from TGF- β 1 deficiency; to rescue the mutant phenotype by TGF- β 1 administration, anti-inflammatory drug therapy or appropriate mouse crosses; to investigate the potential prenatal lethality in homozygous mutant animals; and to investigate expression and function of the pro-region, which was not disrupted by the targeting strategy. TGF- β 1-deficient mice may be valuable models for numerous inflammatory disorders, including autoimmune diseases, transplant rejection, graft versus host reactions, ulcer development, and malignancy-associated cachexia. TGF- β 1-deficient cells are now available for examining the many developmental processes that TGF- β 1 has been reported to affect, including cell proliferation and differentiation, extracellular matrix protein expression, cell adhesion and migration, and angiogenesis. Embryonic and fetal tissue explants from homozygous mutant animals can be used to study *in vitro* organ development in the absence of TGF- β 1. Such studies should further elucidate the multifunctional roles of TGF- β 1 as both a morphogenetic growth factor and immunoregulatory cytokine. □

Received 30 July; accepted 14 September 1992.

- Barnard, J. A., Lyons, R. M. & Moses, H. L. *Biochim. biophys. Acta* **1032**, 79-87 (1990).
- Massague, J. A. *Rev. Cell Biol.* **6**, 597-641 (1990).
- Roberts, A. B. & Sporn, M. B. in *Peptide Growth Factors and Their Receptors I* (eds Sporn, M. B. & Roberts, A. B.) 419-472 (Springer, New York, 1990).
- Heine, U. I. et al. *J. Cell Biol.* **105**, 2861-2876 (1987).
- Lehner, S. A. & Akhurst, R. J. *Development* **104**, 263-273 (1988).
- Wilcox, J. N. & Derynck, R. *Molec. cell. Biol.* **8**, 3415-3422 (1988).
- Gatherer, D., Ten Dijke, P., Baird, D. T. & Akhurst, R. J. *Development* **110**, 445-460 (1990).
- Pelton, R. W., Dickinson, M. E., Moses, H. L. & Hogan, B. L. M. *Development* **110**, 609-620 (1990).
- Schmid, P., Cox, D., Bilbe, G., Maier, R. & McMaster, G. K. *Development* **111**, 117-130 (1991).
- Millan, F. A., Denhez, F., Kondaiah, P. & Akhurst, R. J. *Development* **111**, 131-144 (1991).
- Pelton, R. W., Saxena, B., Jones, M., Moses, H. L. & Gold, L. I. *J. Cell Biol.* **115**, 1091-1105 (1991).
- Thompson, N. L. et al. *J. Cell Biol.* **108**, 661-669 (1989).
- Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. *J. Embryol. exp. Morph.* **87**, 27-45 (1985).
- Bosma, G. C., Custer, R. P. & Bosma, M. J. *Nature* **301**, 527-530 (1983).
- Glode, L. M. & Rosenstreich, D. L. *J. Immun.* **117**, 2061-2066 (1976).
- Kehrl, J. H. et al. *J. Immun.* **137**, 3855-3860 (1986).
- Kehrl, J. H., Thevenin, C., Rieckmann, P. & Fauci, A. S. *J. Immun.* **146**, 4016-4023 (1991).
- Ristow, H. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5531-5534 (1986).
- Wahl, S. M. et al. *J. Immun.* **140**, 3026-3032 (1988).
- Kehrl, J. H. et al. *J. exp. Med.* **163**, 1037-1050 (1986).
- Ortaldo, J. R. et al. *J. Immun.* **146**, 3791-3798 (1991).
- Rook, A. H. et al. *J. Immun.* **136**, 3916-3920 (1986).
- Espevik, T., Figari, I. S., Ranges, G. E. & Palladino, M. A. *J. Immun.* **140**, 2312-2316 (1988).
- Ranges, G. E., Figari, I. S., Espevik, T. & Palladino, M. A. *J. exp. Med.* **166**, 991-998 (1987).
- Wahl, S. M. *J. Clin. Immun.* **12**, 61-74 (1992).
- Gamble, J. R. & Vadas, M. A. *Science* **242**, 97-99 (1988).
- Tsunawaki, S., Sporn, M., Ding, A. & Nathan, C. *Nature* **334**, 260-262 (1988).
- Espevik, T. et al. *J. exp. Med.* **166**, 571-576 (1987).
- Tracey, K. J. et al. *Science* **234**, 470-474 (1986).
- Matthys, P. et al. *Int. J. Cancer* **49**, 77-82 (1991).

- Cronin, M. E., Miller, F. W. & Plotz, P. H. in *Primer on the Rheumatic Diseases* (ed. Schumacher, H. R.) 120-123 (Arthritis Foundation, Atlanta, 1988).
- Talal, N. in *Primer on the Rheumatic Diseases* (ed. Schumacher, H. R.) 136-138 (Arthritis Foundation, Atlanta, 1988).
- Schwartz, R. S. & Datta, S. K. in *Fundamental Immunology* (ed. Paul, W. E.) 819-866 (Raven, New York, 1989).
- Bottazzo, G. F., Pujol-Borrell, R. & Hanafusa, T. *Lancet* **2**, 1115-1118 (1983).
- Londei, M., Lamb, J. R., Bottazzo, G. F. & Feldman, M. *Nature* **312**, 639-641 (1984).
- Brandes, M. E., Allen, J. B., Ogawa, Y. & Wahl, S. M. *J. Clin. Invest.* **87**, 1108-1113 (1991).
- Johns, L. D., Flanders, K. C., Ranges, G. E. & Sriram, S. *J. Immun.* **147**, 1792-1796 (1991).
- Kuruvilla, A. P. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2918-2921 (1991).
- Racke, M. K. et al. *J. Immun.* **146**, 3012-3017 (1991).
- Miller, A., Lider, O., Roberts, A. B., Sporn, M. B. & Weiner, H. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 421-425 (1992).
- Czarniecki, C. W., Chiu, H. H., Wong, G. H. W., McCabe, S. M. & Palladino, M. A. *J. Immun.* **140**, 4217-4223 (1988).
- Brandtzaeg, P. *Curr. Top. Microbiol. Immun.* **146**, 13-25 (1989).
- Coffman, R. L., Leberman, D. A. & Shrader, B. *J. exp. Med.* **170**, 1039-1044 (1989).
- Leberman, D. A., Lee, F. D. & Coffman, R. L. *J. Immun.* **144**, 952-959 (1990).
- Kilshaw, P. J. & Murant, S. J. *Eur. J. Immun.* **21**, 2591-2597 (1991).
- Parker, C. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1924-1928 (1992).
- Chin, Y.-H., Cai, J.-P. & Xiu, X.-M. *J. Immun.* **148**, 1106-1112 (1992).
- Tautz, D. *BioEssays* **14**, 263-266 (1992).
- Frolik, C. A., Dart, L. L., Meyers, C. A., Smith, D. M. & Sporn, M. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3676-3680 (1983).
- Tokuyama, H. & Tokuyama, Y. *Cell Biol. int. Rep.* **13**, 251-258 (1989).
- Cox, D. A. & Burk, R. R. *Eur. J. Biochem.* **197**, 353-358 (1991).
- Thomas, K. R. & Capocchi, M. R. *Cell* **51**, 503-512 (1987).

ACKNOWLEDGEMENTS. We thank G. Michael, E. Choi, P. Bonventre, J. Roths and members of the Doetschman laboratory for discussion; J. Card for photographic assistance; L. Forrester and P. Holm for necropsy and histological techniques; and the Clinical Microbiology Division for analysing the organ cultures. This research was supported by grants from the NIH to T.D. and C.S.

Molecular structure of the acyl-enzyme intermediate in β -lactam hydrolysis at 1.7 Å resolution

Natalie C. J. Strynadka*, Hiroyuki Adachi†, Susan E. Jensen‡, Kathy Johns*, Anita Sielecki*, Christian Betzel§, Kazuo Sutoh† & Michael N. G. James*||

* MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

† Department of Pure and Applied Sciences, College of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan

‡ Department of Microbiology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

§ European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, Hamburg, Germany

|| To whom correspondence should be addressed.

The X-ray crystal structure of the molecular complex of penicillin G with a deacylation-defective mutant of the RTEM-1 β -lactamase from *Escherichia coli* shows how these antibiotics are recognized and destroyed. Penicillin G is covalently bound to Ser 70 O γ as an acyl-enzyme intermediate. The deduced catalytic mechanism uses Ser 70 O γ as the attacking nucleophile during acylation. Lys 73 N ϵ acts as a general base in abstracting a proton from Ser 70 and transferring it to the thiazolidine ring nitrogen atom via Ser 130 O γ . Deacylation is accomplished by nucleophilic attack on the penicilloyl carbonyl carbon by a water molecule assisted by the general base, Glu 166.

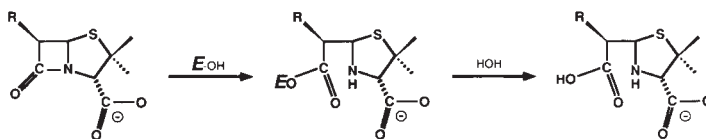
β -LACTAM antibiotics (the penicillins and the cephalosporins) are used as effective, substrate-analogue-based drugs targeted against the bacterial cell-wall synthetic enzymes, peptidoglycan transpeptidases and D-alanyl-D-alanine carboxypeptidases¹⁻³. β -Lactamases constitute a group of bacterial enzymes that destroy these clinically useful antibiotics by cleaving and hydrolysing the sensitive four-membered β -lactam ring (Fig. 1). The class A β -lactamases are serine hydrolases produced by both Gram-positive and Gram-negative bacteria. Crystal structures of the Gram-positive β -lactamases from *Staphylococcus aureus*^{4,5} and from *Bacillus licheniformis*^{6,7} are known at 2.0 Å resolution. Recently, a report on the 2.5 Å resolution crystal structure of the Gram-negative β -lactamase, RTEM-1 from *Escherichia coli* has been published⁸. A wealth of kinetic data is available for many of the class A β -lactamases and their site-directed mutants⁹⁻²². Despite all of this information there seems to be no consensus regarding the nature of the general base involved in the acylation step of the reaction^{1,3}. In addition, various different roles have been assigned to some of the conserved residues within the active sites. The results discussed here are based on our independent determinations of the crystal structures of RTEM-1 from *E. coli* at 1.9 Å resolution, the mutant Glu 166 Asn (E166N) RTEM-1 and its complex with benzyl penicillin (Pen G) at 1.7 Å resolution. The E166N mutant was shown to be defective in deacylation¹². The present crystallographic analysis shows that we have trapped the natural sub-

strate, Pen G, covalently linked to Ser 70 O γ as an acyl-enzyme complex. Comparison of these molecular structures allows us to propose a revised catalytic mechanism for the class A β -lactamases that is consistent with the structural, chemical and kinetic data presently available.

Native RTEM-1

The plasmid-encoded class A β -lactamase RTEM-1 is found widely disseminated in *E. coli* strains and other Enterobacteriaceae^{23,24}. We have crystallized RTEM-1 and determined its three-dimensional structure by a combination of multiple isomorphous replacement²⁵ and molecular replacement²⁶ techniques. A view of the current $2||F_o| - |F_c||$ electron-density map with the refined molecular model in the vicinity of the active site is shown in Fig. 2a. Comparison of the RTEM-1 molecular structure with those of the β -lactamases of *S. aureus*^{4,5} and of *B. licheniformis*^{6,7} available from the Protein Data Bank²⁷ indicates that the three molecules are topographically similar. Nonetheless, the root-mean-square (r.m.s.) differences of the 254 common C α coordinates in the pairwise comparisons of RTEM-1 against *S. aureus* and RTEM-1 against *B. licheniformis* are relatively large (2.4 Å and 2.1 Å, respectively), owing to many regions of common secondary structural units (α -helices and strands of β -sheets) having different relative orientations. There are also large variations in the conformations of some of the loop regions in order to accommodate small insertions and/or

FIG. 1 The two steps in the β -lactamase-catalysed hydrolysis of benzyl penicillin (Pen G: R = NHCCH₂C₆H₅). The acylation step is the result of nucleophilic attack of Ser 70 O γ on the carbonyl-carbon atom (C7) of the β -lactam ring¹. Electrophilic assistance for this step is provided by an oxyanion 'hole' in an analogous way as in the serine proteinases^{29,30,35}. The presence of the acyl-enzyme intermediate has been firmly established by a number of detailed kinetic studies^{17,19,22}. The deacylation step is broadly accepted to be a general-base-assisted nucleophilic attack by a water molecule on the carbonyl-carbon atom of the ester bond to Pen G. The general base in the deacylation step is the highly conserved side-chain of Glu 166 as originally suggested by the structural studies on the *S. aureus* enzyme⁴ and subsequently confirmed by several site-directed mutagenesis studies¹¹⁻¹³ that resulted in deacylation-defective enzymes.

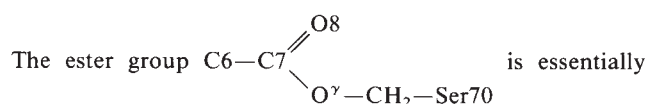


RTEM-1 the distance from Lys 73 N^ϵ to Ser 130 O^γ is 3.7 Å, far too long for a strong hydrogen-bonding interaction. Ser 130 O^γ forms a strong hydrogen bond to Lys 234 N^ϵ (2.8 Å) and to a water molecule WAT 454 (2.8 Å). Lys 234 N^ϵ forms two additional hydrogen bonds to Ser 235 O (3.1 Å) and to WAT 430 (3.1 Å).

The Glu 166 side chain lies at the bottom of the active site depression where the carboxylate group forms a hydrogen bond to WAT 321 (2.7 Å) and to the side-chain amide nitrogen of Asn 170 (3.1 Å). Water WAT 321 makes an additional hydrogen bond to the sidechain amide of Asn 170 $O^{\delta 1}$ (2.5 Å). This solvent is also observed in the structures of the related class A β -lactamases⁴⁻⁷ and has been implicated as the deacylating water molecule^{1,4}.

Acyl-enzyme intermediate

Deacylation-defective mutants of β -lactamases have been made by site-specific mutagenesis: *Bacillus cereus* E166D¹¹; RTEM-1, E166A, E166Q, E166D and E166N¹²; *B. licheniformis* E166A¹³. With the RTEM-1 mutants the acylenzyme accumulates, clearly demonstrating the importance of Glu 166 in deacylation and proving that deacylation is not simply the reverse of acylation with water acting as the nucleophile¹². We have crystallized the E166N mutant of RTEM-1 and have solved and refined that structure at 1.7 Å resolution (Fig. 3). A difference electron-density map computed with coefficients representing the differences in structure-factor amplitudes from that crystal and from a second one soaked in a solution containing Pen G reveals electron density accounting for all of the atoms of Pen G covalently bonded from the carbonyl-carbon atom C7 of the antibiotic to Ser 70 O^γ . These two structures have been refined sufficiently to define the Pen G geometry and its stabilizing hydrogen-bonding interactions (legend to Fig. 3).



In addition to the covalent bond from C7 to Ser 70 O^γ there are other stabilizing hydrogen-bonding interactions and one salt bridge (Fig. 4). As predicted from earlier modelling studies in the class A β -lactamases^{4,6} the carbonyl-oxygen atom O8 of the complex forms two strong hydrogen bonds to the main-chain nitrogen atoms of Ser 70 (2.7 Å) and Ala 237 (3.0 Å). These interactions comprise the oxyanion binding site of the enzyme (Fig. 4). Very similar distances (2.7 Å and 2.9 Å, respectively) were reported for the interaction of an acyl-ester carbonyl-oxygen atom to the oxyanion hole nitrogens in the 2.5 Å resolution structure of a monobactam (aztreonam)-inhibited form of the class C β -lactamase from *Citrobacter freundii*³³. Atoms N14 and O16 (see Fig. 3 for atom numbering), which form the R group peptide on Pen G, make hydrogen bonds with the main-chain carbonyl-oxygen atom of Ala 237 (2.9 Å) and with the amide-nitrogen atom of Asn 132 $N^{\delta 2}$ (2.6 Å), respectively. The Pen G carboxylate oxygen atoms O12 and O13 form strong hydrogen-bonds to the side-chain nitrogen atom of Arg 244 $N^{\eta 1}$ (2.7 Å) and to the side-chain oxygen atom of Ser 235 O^γ (2.7 Å). This is consistent with site-specific mutagenesis experiments which indicated that Arg 244 contributed ~1.3–2.3 kcal mol⁻¹

FIG. 3 *a*, The difference electron-density map (coefficients $|F_{\text{complex}}| - |F_{\text{mutant}}|$), and calculated phases for the partially refined E166N mutant structure, $R=0.22$) contoured at $0.20 \text{ e}\text{\AA}^{-3}$ in the region of the active site (where $|F_{\text{complex}}|$ and $|F_{\text{mutant}}|$ are the structure factor amplitudes of the mutant crystal soaked in solutions containing Pen G and the native mutant crystal, respectively). The map clearly shows the cleaved Pen G molecule (thick lines). *b*, A schematic illustration of benzyl penicillin (pen G) with atom numbering as referred to in the text.

METHODS. The E166N mutant of RTEM-1 β -lactamase was overproduced in *E. coli* MV1184 carrying the pHA508-166N plasmid¹². Isomorphous crystals of unit cell dimensions $a=62.2$, $b=88.3$, $c=41.6$ Å were grown as for the native enzyme and using native seed crystals for the initial nucleation. The cocrystals of the E166N mutant RTEM-1 and penicillin G were obtained by soaking crystals in 10 molar excesses of the antibiotic for 10 hours before data collection. The soaking solution was replaced with fresh antibiotic every 2 h. Data to 1.7 Å resolution were collected for the E166N crystals in the absence and presence of pen G using synchrotron radiation as outlined above for native RTEM-1. The least-squares program PROLSQ was used to refine the native RTEM-1 coordinates (replacing E166 with an asparagine side chain) with the uncomplexed E166N intensity data (24,067 observations between 8 and 1.7 Å resolution). A difference Fourier map was then computed using coefficients $|F_{\text{complex}}| - |F_{\text{mutant}}|$, and phases calculated from the coordinates of the partially refined E166N mutant structure. Both the uncomplexed and complexed E166N models have been further refined to present R factors of 0.184 and 0.182, r.m.s. deviations in bond distances 0.014 Å and 0.016 Å, bond angles 3.1° and 3.1°, and 120 and 114 currently identified waters, respectively for all data from 10–1.7 Å resolution with $|F| > 1\sigma|F|$. Refined models will be reported in detail elsewhere and the coordinates submitted to the Brookhaven Protein Data Bank on completion of the refinement.

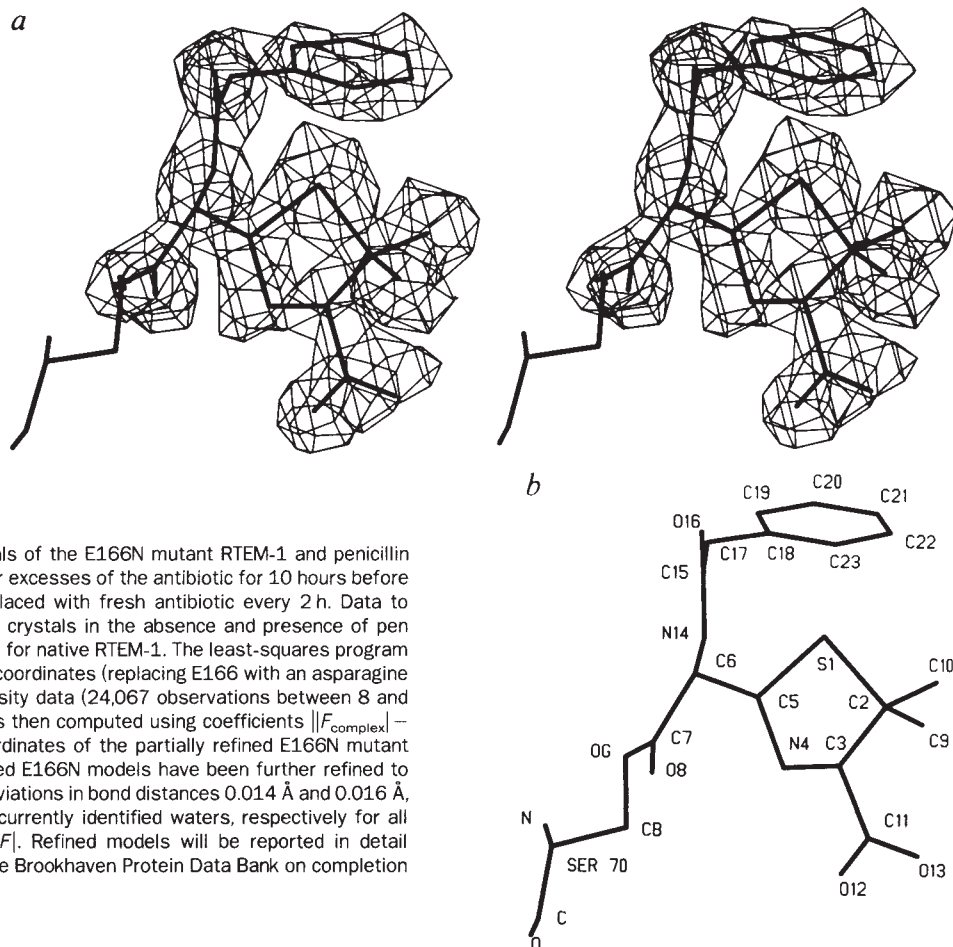
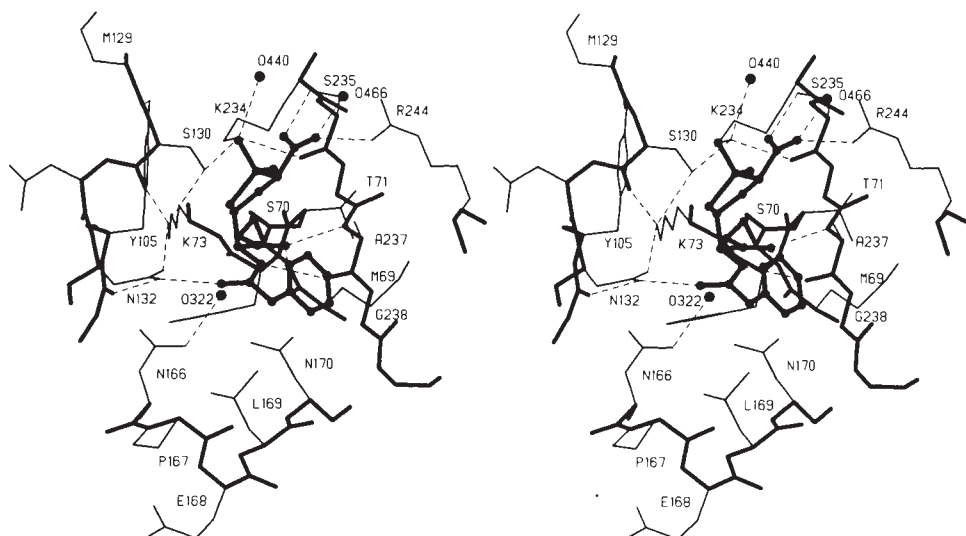


FIG. 4 A stereographic view of penicillin G covalently bound to Ser 70 O^γ in the active site of RTEM-1 β -lactamase. The bonds between the atoms of Pen G and those of the main chain of RTEM-1 are drawn with thick lines; side-chain atoms are connected by thin lines and hydrogen bonds are represented as thin dashed lines. WAT 322 is equivalent to the deacylating WAT321 in wild-type RTEM-1 but displaced by 1.1 Å in the E166N mutant (see Fig. 2b).



to the binding of penicillins both in the ground state and the transition state species of RTEM-1 β -lactamase³⁴. There is no direct hydrogen-bond from either Ser 130 or Lys 234 to the C3 carboxylate moiety as deduced from previous modelling studies^{4,6}. The closest distance between Lys 234 N^ε and the C3 carboxylate is 3.5 Å, suggesting only a weak electrostatic stabilization of the substrate by Lys 234. Finally, in the observed, cleaved form, the thiazolidine ring nitrogen N4 is 3.1 Å from the side-chain oxygen of Ser 130. Although this distance is acceptable, the two groups are poorly oriented to support a hydrogen-bond in which N4 would be the donor atom (Fig. 4). The interaction between Ser 130 O^γ and N4 in the acyl-enzyme intermediate can therefore be described as a normal van der Waals contact.

Generally, the hydrogen-bonding interactions among the amino-acid residues constituting the active site of the RTEM-1 mutant are maintained in the native and in the acyl-enzyme complex (see Figs 2b and 4). Lys 73 N^ε forms a strong hydrogen-bond with Asn 132 O^{δ1} and with Ser 130 O in both structures. Lys 234 N^ε forms the same constellation of three hydrogen-bonds to Ser 130 O^γ, Ser 235 O and WAT 430(440) in both structures as well. The relatively long non-bonded contact distance between Ser 70 O^γ and Ser 130 O^γ (3.3 Å) is also maintained both in the wild-type and the acyl-enzyme structures. There is one significant difference, however, in the hydrogen-bonding interactions involving the triad Ser 70, Lys 73 and Ser 130. In the wild-type RTEM-1 enzyme, Lys 73 N^ε is hydrogen-bonded to Ser 70 O^γ (2.7 Å) but not to Ser 130 O^γ (3.7 Å). In the acyl-enzyme complex, slight conformational changes in both Lys 73 and Ser 130 reverse the situation so that the distance from Lys 73 N^ε to Ser 130 O^γ is reduced to 2.8 Å and these residues share a hydrogen-bond. In contrast, the distance from Lys 73 N^ε to Ser 70 O^γ is increased to 3.3 Å in the acyl-enzyme complex and these residues no longer share a hydrogen-bond (Fig. 4).

It has been assumed that Lys 73 N^ε is fully protonated and thus positively charged^{4,5}. In this context, the role of Lys 73 would be to aid the direct transfer of the serine 70 proton to the leaving-group nitrogen by providing a suitable electrostatic potential gradient¹. On the basis of the structural results on the acyl enzyme intermediate with Pen G, we propose rather, that Lys 73 could have a depressed proton affinity (pK_a) and a neutral side-chain amino-group. In this form it would be ready to act as a general base assisting in the nucleophilic attack of Ser 70 O^γ on the carbonyl-carbon of the substrate.

A new view of the catalytic mechanism

Figure 5 is a schematic representation of our interpretation of the chemical events involved in class A β -lactamase-catalysed

hydrolysis of β -lactam antibiotics. The scheme is based on the crystal structural data discussed above for native RTEM-1 and the penicilloyl complex of the E166N mutant. The model of the Michaelis complex is derived from the observed binding interactions in the acyl-enzyme intermediate (Figs 4 and 5a). Hydrogen-bonding interactions that would stabilize the initial binding of Pen G to RTEM-1 include: the oxyanion hole (Ser 70 NH and Ala 237 NH), the Pen G side-chain peptide to Ala 237 O and Asn 132 N^{δ2}, the electrostatic interactions of the C3 carboxylate with Arg 244 N^{η1}, Ser 235 O^γ and Lys 234 and, based on the distance observed in the acyl-enzyme, a proposed hydrogen-bond from Ser 130 O^γ to the thiazolidine ring atom N4 (the leaving-group nitrogen atom).

Catalysis is initiated by nucleophilic attack of Ser 70 O^γ on the *re*-face of the β -lactam ring. Electrophilic assistance is provided by the hydrogen-bonding interactions from the oxyanion hole to the carbonyl-oxygen atom O8. This would enhance the polarization of this bond, aiding the attack of Ser 70 O^γ on the electrophilic carbonyl-carbon atom, C7. We propose that Lys 73 is unprotonated in the native enzyme and in the Michaelis complex. It thus assists in the nucleophilic attack by acting as a general base thereby accepting the proton from Ser 70 O^γ as the first tetrahedral intermediate is formed. As mentioned, a strong hydrogen bond between these groups pre-exists in the native enzyme, thus defining the proton transfer pathway. The unprotonated Lys 73 amino-group is chemically equivalent to the general-base character of the imidazole ring of His 57 in the serine proteinase mechanism³⁵. The positional equivalence of these two groups was shown by superimposing the active site residues in the bacterial serine proteinase SGPB with *S. aureus* β -lactamase⁴. Earlier work had suggested that other residues could act as the general base in the acylation step (Ser 130 (ref. 33) or Glu 166 (ref. 11, 36)). The long distances from Ser 70 to these other groups in the class A β -lactamases clearly precludes direct proton transfer to them. Thus Lys 73 is the much more likely candidate for the general base.

We also propose that the conformational changes involving the side chains of Lys 73 and Ser 130 (Figs 2b and 4) would take place after the first proton transfer from Ser 70 O^γ to Lys 73 N^ε. These changes develop the second proton-transfer pathway from Lys 73 N^ε to the nitrogen of the leaving group by means of Ser 130 O^γ (Fig. 5b). The second transfer involves two protons, one from Lys 73 N^ε to Ser 130 O^γ and the other from Ser 130 O^γ to N4 of the thiazolidine ring. In the uncleaved form of Pen G and in β -lactams in general, the nitrogen atom is already pyramidal^{37,38}. In the Michaelis complex the nitrogen lone-pair electrons would be directed towards Ser 130 O^γ, ready to accept the proton, as suggested above and deduced from the acceptable hydrogen-bond distance observed in the acyl-enzyme

structure. Thus, the formation and collapse of tetrahedral intermediate I (Fig. 5a–c) involves three protons in a shuttle from Ser 70 O^γ to Lys 73 N^ε to Ser 130 O^γ and finally to the nitrogen of the leaving group. Site-directed mutagenesis experiments on β -lactamases have shown that substitution of either Lys 73 by arginine¹¹ or of Ser 130 by alanine or glycine²¹ leads to severe impairment of the acylation step of the reaction, confirming the importance of these residues in the mechanism. In the *Streptomyces albus* G β -lactamase the S130G mutant exhibits ~fivefold more activity towards a variety of substrates than the S130A mutant²¹. It is possible that in the S130G mutant a water molecule could perform the role equivalent to the serine in the

proton-transfer step (Fig. 5b).

The above proposal implies that the relative proton affinities (pK_a s) of both Lys 73 and Ser 130 side chains would have to be significantly reduced to aid these transfers efficiently. The pH rate profiles for the related RTEM-2 β -lactamase indicate the presence of two acid-dissociable groups with pK_a s of 6.2 and 7.6 in the vicinity of Ser 70 (ref. 36). The assignment of these pK_a s to specific residues in the active site has not been confirmed. The electrostatic environment of the Lys 73 side-chain nitrogen is such that it is not easy to predict its proton affinity. Site-directed mutations of Asn 132 to either an alanine or a serine in *S. albus* G class A β -lactamase resulted in a

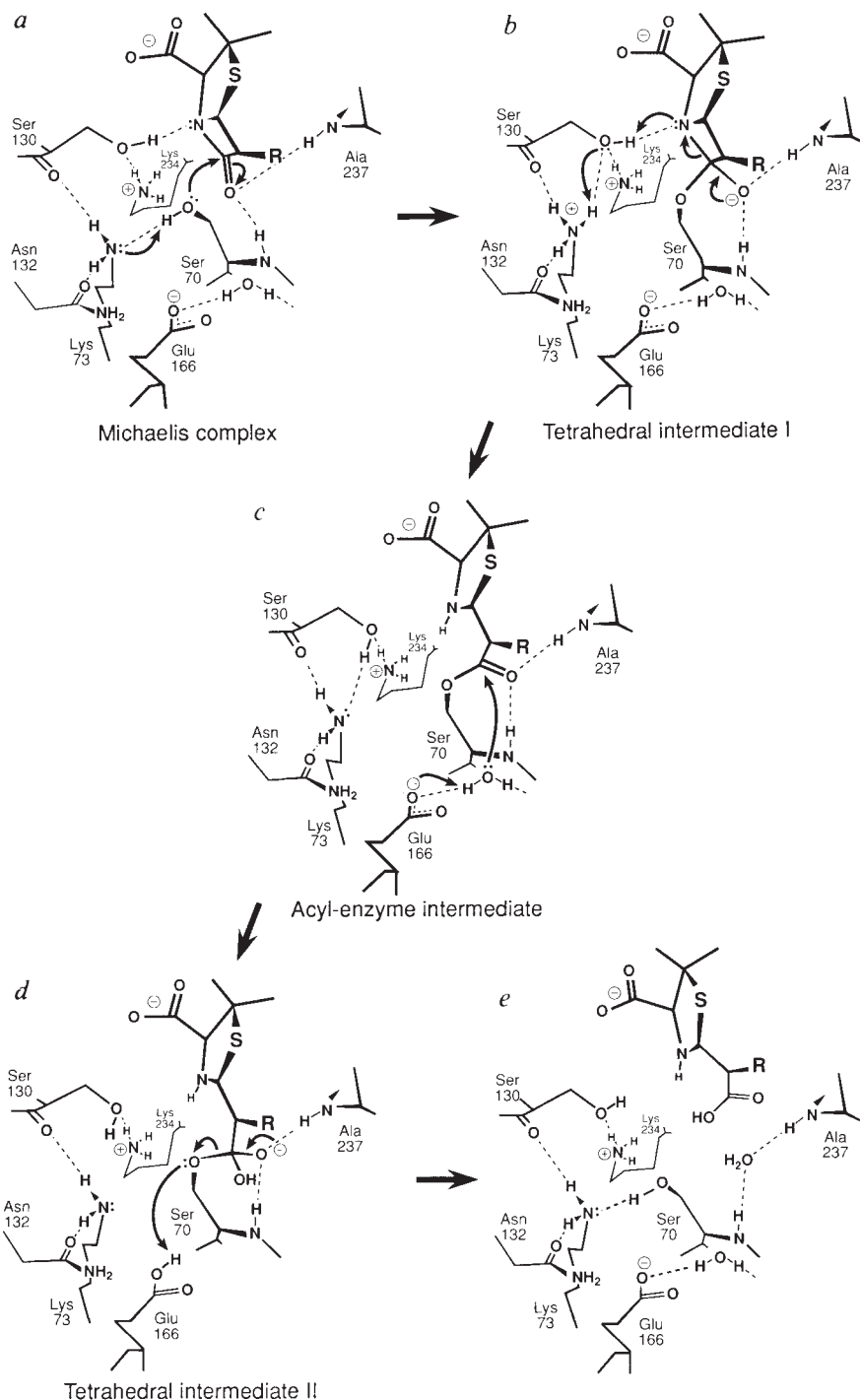


FIG. 5 The proposed mechanism of β -lactam hydrolysis by class A β -lactamases as deduced from the 1.7 Å refined structure of the acyl-enzyme intermediate.

dramatic reduction of the specific rate constant k_{cat} and a conservative increase in the Michaelis constant K_{m} such that $k_{\text{cat}}/K_{\text{m}}$ for these mutants was only 0.5% that of wild-type enzyme for the best substrate^{20,21}. This could be explained by the loss of the hydrogen-bond acceptor function on mutation of Asn 132 that perturbs the stereochemical environment of the Lys 73 general base and affects the acylation rate. The effect on K_{m} could be attributed to the loss of hydrogen bonding for the substrate peptide carbonyl O16 (Fig. 4).

It has been pointed out that adjacent positive charges can cause large decreases in the pK_{a} of amino-groups in proteins³³. Thus, the positively charged side chain of Lys 234 at a distance of 5.3 Å in wild-type and 5.0 Å in the complex, could assist in lowering the pK_{a} of Lys 73. Following the suggestion of Oefner *et al* for the class C enzymes³³, we also suggest that the pK_{a} of Ser 130 could be reduced significantly by the direct hydrogen-bonded interaction with the two positively charged side chains of Lys 73 and Lys 234 (2.7 Å and 2.8 Å, respectively in the acyl-enzyme). Indeed, kinetic analyses of Lys 234 mutants led to the conclusion that this residue did not contribute a significant electrostatic interaction to substrate in the ground state but it was important in optimizing the electrostatic potential of the transition state during acylation¹⁴. At pH values less than 6.0 the substitution of Lys 234 by alanine or by glutamate had little effect on binding (K_{m}) but both mutants had decreased k_{cat} of up to 800-fold¹⁴. These results are consistent with our observation that Lys 234 forms only a relatively weak electrostatic interaction with the C3 carboxylate of Pen G in the acyl-enzyme intermediate. The strong hydrogen-bond to Ser 130 O γ from Lys 234 N ϵ is crucial in lowering the proton affinity of the serine hydroxyl group thereby aiding proton transfer.

The deacylation steps in the hydrolytic reaction catalysed by

class A β -lactamases are shown in Fig. 5c-e. As many before have proposed^{1,4,12,13}, deacylation is accomplished by a general-base-assisted nucleophilic attack of the deacylating water (WAT 321, Fig. 2b) on the ester carbonyl carbon of the acyl-enzyme intermediate (Fig. 5c). This results in tetrahedral intermediate II (Fig. 5d). The ester carbonyl oxygen of the acyl-enzyme intermediate is bound in the oxyanion hole thus creating a favourable electrophilic assistance to the OH $^-$ attack. The general base is Glu 166 carboxylate which accepts the proton from the deacylating water molecule and transfers it to Ser 70 O γ in order to regenerate the active site (Fig. 5e). We show in Fig. 5d that this transfer is direct from Glu 166 to Ser 70. It is also possible that Lys 73 could assist in the deacylation steps by providing a means to shuttle the proton from Glu 166 to Ser 70. Small conformational changes in Lys 73 and Glu 166 could establish this proton transfer pathway. The kinetic analysis of the K73R mutant of the *B. cereus* β -lactamase showed that the deacylation rate is also altered by the replacement of the Lys 73 side chain¹¹.

The observed accumulation of the acyl-enzyme intermediate in the E166N mutant of RTEM-1¹², and the observed crystal structure of its complex presented here are a direct confirmation of this segment of the pathway. Why is the E166N mutant deacylation defective? The answer clearly points to the loss of the general base carboxylate function of Glu 166 in the mutant enzyme. But the deacylating water is still present in the acyl-enzyme complex (Fig. 4). It is moved by ~ 1.1 Å from its presumably ideal location as observed in the wild-type RTEM-1 (Fig. 2b) so that the direction of nucleophilic attack is also impaired. The E166N mutant does deacylate slowly¹². This could be accomplished by this same water (WAT 321) but it is certainly not as effective a nucleophile as in the wild-type RTEM-1. $\square$

Received 11 August; accepted 28 September 1992.

- Herzberg, O. & Moulton, J. *Curr. Opin. Struct. Biol.* **1**, 946-953 (1991).
- Frère, J.-M. & Joris, B. *Crit. Rev. Microbiol.* **11**, 299-396 (1985).
- Ghuysen, J.-M. *A. Rev. Microbiol.* **45**, 37-67 (1991).
- Herzberg, O. & Moulton, J. *Science* **236**, 694-701 (1987).
- Herzberg, O. *J. molec. Biol.* **217**, 701-719 (1991).
- Moews, P. C., Knox, J. R., Dideberg, O., Charlier, P. & Frère, J.-M. *Proteins Struct. Funct. Genet.* **7**, 156-171 (1990).
- Knox, J. R. & Moews, P. C. *J. molec. Biol.* **220**, 435-455 (1991).
- Jelsch, C., Lenfant, F., Masson, J. M. & Samama, J. P. *FEBS Lett* **299**, 135-142 (1992).
- Knowles, J. R. *Acc. Chem. Res.* **18**, 97-104 (1985).
- Christensen, H., Martin, M. T. & Waley, S. G. *Biochem. J.* **266**, 853-861 (1990).
- Gibson, R. M., Christensen, H. & Waley, S. G. *Biochem. J.* **272**, 613-619 (1990).
- Adachi, H., Ohta, T. & Matsuzawa, H. *J. Biol. Chem.* **266**, 3186-3191 (1991).
- Escobar, W. A., Tan, A. K. & Fink, A. L. *Biochemistry* **30**, 10783-10787 (1991).
- Ellerby, L. M., Escobar, W. A., Fink, A. L., Mitchinson, C. & Wells, J. A. *Biochemistry* **29**, 5797-5800 (1990).
- Dalbadi-McFarland, G., Neitzel, J. J. & Richards, J. H. *Biochemistry* **25**, 332-338 (1986).
- Martin, M. T. & Waley, S. G. *Biochem. J.* **254**, 923-925 (1988).
- Virden, R., Tan, A. K. & Fink, A. L. *Biochemistry* **29**, 145-153 (1990).
- Healey, W. J., Labgold, M. R. & Richards, J. H. *Proteins Struct. Funct. Genet.* **6**, 275-283 (1989).
- Fisher, J., Belsaco, J. G., Khosla, S. & Knowles, J. R. *Biochemistry* **19**, 2895-2901 (1980).
- Jacob, F. *et al. Protein Engng* **4**, 79-86 (1990).
- Jacob, F., Joris, B., Lepage, S., Dusart, J. & Frère, J.-M. *Biochem. J.* **271**, 399-406 (1990).
- Cartwright, S. J., Tan, A. K. & Fink, A. L. *Biochem. J.* **263**, 905-912 (1989).
- Abraham, E. P. & Chain, E. *Nature* **146**, 837 (1940).
- Matthew, M. & Hedges, R. W. *J. Bact.* **125**, 713-718 (1976).
- Green, D. W., Ingram, V. M. & Perutz, M. F. *Proc. R. Soc. A* **A225**, 287-307 (1954).
- Rossmann, M. G. *The Molecular Replacement Method*. Int. Sci. Rev. Vol. 13 (Gordon & Breach, New York, 1972).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535-542 (1977).

- Ambler, R. P. *et al. Biochem. J.* **276**, 269-272 (1991).
- Henderson, R. *J. molec. Biol.* **54**, 341-354 (1970).
- Robertus, J. D., Kraut, J., Alden, R. A. & Birktoft, J. J. *Biochemistry* **11**, 4293-4303 (1972).
- Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A. in *The Chemistry of Penicillin* (eds Clarke, H. T., Johnson, J. R. & Robinson, R.) 310-367 (Princeton Univ. Press, 1949).
- Dexter, D. D. & van der Veen, J. M. *J. chem. Soc.* **185**, 110-115 (1978).
- Oefner, G. *et al. Nature* **343**, 284-288 (1990).
- Zafaralla, G., Manavathu, E. K., Lerner, S. A. & Mobashery, S. *Biochemistry* **31**, 3847-3852 (1992).
- Kraut, J. *A. Rev. Biochem.* **46**, 331-358 (1977).
- Knap, A. K. & Pratt, R. F. *Biochem. J.* **273**, 85-91 (1991).
- James, M. N. G., Hall, D. & Hodgkin, D. C. *Nature* **220**, 168-170 (1968).
- Sweet, R. M. & Dahl, L. F. *J. Am. chem. Soc.* **92**, 5489-5507 (1970).
- Viera, J. & Messing, J. *Meth. Enzym.* **153**, 3-11 (1987).
- Hamlin, R. *Meth. Enzym.* **114**, 416-452 (1985).
- Howard, A. J., Nielson, C. & Xuong, N. H. *Meth. Enzym.* **114**, 452-472 (1985).
- Otinowski, Z. *Am. Crystallogr. Ass. Ann. Meeting*, New Orleans, Abstr. C04 (1990).
- Navaza, J. *Acta crystallogr.* **A46**, 619-620 (1990).
- Read, R. J. & Schierbeek, A. J. *J. appl. Crystallogr.* **21**, 490-495 (1988).
- Read, R. J. *Acta crystallogr.* **A42**, 140-149 (1986).
- Jones, T. A. *J. Appl. Crystallogr.* **11**, 268-272 (1978).
- Leslie, A. G. W., Brick, P. & Wonacott, A. J. *CCP4 Newsletter* **18**, 33-39 (1986).
- Brünger, A. J. *J. molec. Biol.* **203**, 803-816 (1988).
- Hendrickson, W. A. & Konner, J. H. in *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. & Venkatesan, K.) 1301-1323 (Indian Institute of Sciences, Bangalore 1980).

ACKNOWLEDGEMENTS. We thank K. Wilson, Z. Dauter and the staff at the Hamburg Outstation EMBL, for support for the data collection and M. Wylie for handling the manuscript during preparation. The research in Edmonton was funded by the Medical Research Council of Canada in a grant to the Group in Protein Structure and Function at the University of Alberta. The research in Tokyo was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan to H.A. N.C.J.S. is the recipient of the J. Gordin Kaplan Memorial Postdoctoral Fellowship at the University of Alberta.

Spin-up and recovery in the 1989 glitch of the Crab pulsar

A. G. Lyne, F. Graham Smith & R. S. Pritchard

Nuffield Radio Astronomy Laboratories, University of Manchester,
Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

THE rotation of the Crab pulsar, as measured by pulsed radio and optical emission, has been monitored almost continuously since its discovery in 1968. A steady decrease in the rotation rate has been interrupted at intervals of about five years by glitches: discontinuous increases in rotation speed followed by partial recovery in the form of an exponential return to a new slowdown rate on a timescale of about 20 days. The recovery is usually interpreted as the re-establishment of stable differential rotation between the neutron star crust and part of its superfluid interior. Here we describe the largest glitch so far recorded in the Crab pulsar¹. It occurred in 1989 while observations were being made, and the recovery was followed in unprecedented detail. The spin-up itself was also partly resolved in time, and the whole event, glitch plus recovery, is now seen to include three distinct exponential components, with timescales of 1, 20 and 300 days. The physical interpretation of these distinct components is unclear. The 1989 glitch, like the 1975 one², caused a permanent increase in the slowdown rate, of a magnitude that is difficult to reconcile with current neutron star models.

The 1989 glitch occurred at 09:36 UT on 29 August during a long series of 610-MHz radio observations at Jodrell Bank³. This 11-year series combined with earlier observations of the pulsar gives us a fairly complete account of the rotational behaviour of this pulsar over a period of 23 years, including the glitches of 1969, 1975 and 1986 (refs 4–6).

The pulse arrival times at the observatory were corrected for the Earth's rotation and orbital motion using the Jet Propulsion Laboratory barycentric ephemeris DE 200 (ref. 7) to refer them to the Solar System barycentre. Barycentric rotational frequencies were obtained by fitting to a set of corrected arrival times, typically over intervals of ~6 h close to the glitch and 50 days away from it. The glitch itself is a small perturbation on the normal slowdown of rotation, whose character may be deter-

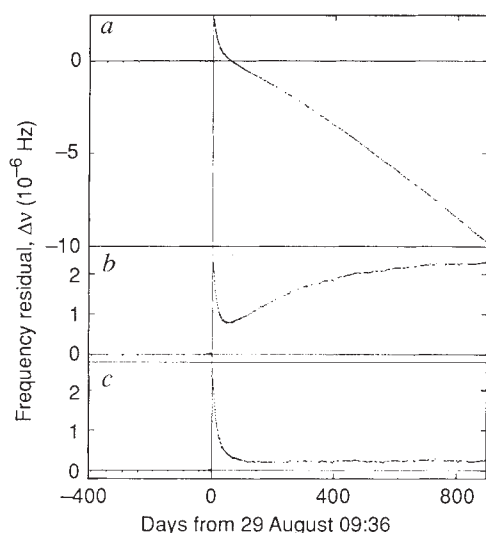


FIG. 1 The 1989 glitch in the Crab pulsar. Daily values of residuals in (a) the rotation rate ν , (b) the rotation rate after subtracting the change of slope $\Delta\dot{\nu}$ and (c) after also subtracting the 265-day asymptotic exponential. All rotation rates have been corrected for the Earth's rotation and orbital motion, and are referred to a best fit for rotation rate and its first derivative made over 400 days immediately preceding the glitch.

TABLE 1 Components of the 1989 glitch in the Crab pulsar

	Amplitude	Characteristic time (days)
Initial unresolved step	1.85×10^{-6} Hz	<0.1
Asymptotic exponential	0.7×10^{-6} Hz	0.8
Exponential decay	2.3×10^{-6} Hz	18
Asymptotic exponential	2.1×10^{-6} Hz	265
Change in derivative	$-155 \times 10^{-15} \text{ s}^{-2}$	

mined from our long series of data. The normal slowdown is represented by

$$\nu(t) = \nu(t_0) + \dot{\nu}(t - t_0) + \frac{1}{2}\ddot{\nu}(t - t_0)^2$$

The parameters of this normal pre-glitch behaviour were obtained from a best fit for rotation rate ν and its first derivative $\dot{\nu}$, made over the preceding 400 days; the value of the second derivative $\ddot{\nu}$ was obtained from the full data set from 1969 to 1992. The plot of rotation rates in Fig. 1a shows the residuals relative to this normal slowdown.

The immediate effect of the glitch is a step increase in rotation speed; in this respect the glitch is similar to those observed in other pulsars such as the Vela pulsar⁸. Remarkably, the subsequent behaviour leaves the pulsar rotating more slowly after 50 days than it would have done in the absence of the glitch. This is because on a long timescale the main effect of the glitch is an increase in slowdown rate, which is responsible for the slope in the later part of Fig. 1a. The slowdown rate can be seen to approach a constant value after ~500 days. We now show that there are also three transient components on timescales of ~300 days, 20 days and 1 day; of these, only the second has previously been reported.

Figure 1b shows the rotation rate residuals after allowing for the increased value of slowdown rate which fits the later part of Fig. 1a. The two longer transients are now seen, one as an exponential decay and the other as an asymptotic exponential rise with 265-day time constant. Figure 1c shows the residuals after removing the 265-day component. Except for the first two days, the decaying component is well fitted by an exponential with 18-day time constant, but we now see that the glitch also leaves a nondecaying increase in rotation rate.

Finally, in Fig. 2a we plot with higher time resolution the residuals obtained from all our observations in the first 15 days after the glitch; as in Fig. 1c, the effects of the increased slowdown rate and the 265-day component have been removed. The onset of the glitch is now seen to comprise an unresolved step followed by an exponential asymptotic rise to meet the 18-day exponential decaying component. The time constant of this asymptotic rise is 0.8 day.

The amplitude of the initial unresolved step is 1.85×10^{-6} Hz, and that of the short-term asymptotic rise is 0.7×10^{-6} Hz. The amplitude of the decaying exponential (2.3×10^{-6} Hz) is less than the sum of these, leaving a net increase of 0.25×10^{-6} Hz. The long-term asymptotic rise eventually adds a further increase of 2.1×10^{-6} Hz to the rotation rate.

The amplitudes and time constants of these components of the glitch are given in Table 1. The values, obtained by a best-fitting process, are accurate to ~10%. They are, however, not completely independent of each other; for example, satisfactory fits can be obtained with time constants between 16 days and 20 days for the main recovery component, with small adjustments to the other values. Other equally valid fits might also be obtained by allowing the 18-day decay to depart from a pure exponential; it would then be possible, for example, to obtain a solution in which the two asymptotic increases account for the whole of the increase in rotation rate.

Delayed increases in rotation rate have not previously been observed in glitches in any pulsar, on either a short or a long

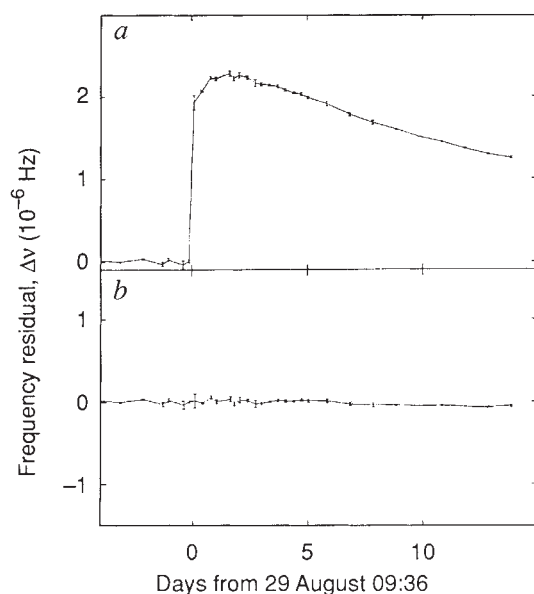


FIG. 2 *a*, Observed values of rotation rate for the first 15 days; as in Fig. 1c, all values are residuals after allowing for the normal slowdown rate, the persistent increase in slowdown rate, and the 265-day component. *b*, The residuals after subtracting all four components described in the text.

timescale. It is remarkable that simple exponentials provide good fits to all three time-varying components, indicating that three distinct physical processes, possibly in three distinct regions, occurred in the superfluid of the neutron star. The exponential decay is regarded as the re-establishment of the normal differential rotation between the crust and the superfluid; the two delayed spin-ups have as yet no explanation.

The largest and most significant effect of the glitch is seen to be the persistent change in slowdown rate, amounting to 0.04%.

It has already been pointed out² that the 1975 glitch left a persistent increase of $\sim 0.05\%$ in slowdown rate; the available data between 1975 and 1989 show no indication of a decay in this increase. The change in 1989 was 0.04%, so that the cumulative effect of the two glitches is an increase in slowdown rate of $\sim 0.1\%$ over 20 years.

The rotational slowdown rate is determined by the torque due to electromagnetic radiation and particle outflow, and by the effective moment of inertia of the star. We are not yet able to say which of these two parameters changed at the glitches. It is difficult to account for so large a change in terms of a decrease in effective moment of inertia due to a change in the shape of the crust of the star, because for a rotation rate of 30 Hz the equilibrium ellipsoid has a moment of inertia only 10^{-4} greater than that of a sphere. Alternatively the decrease might be due to a progressive increase in vortex pinning in the superfluid as the neutron star slows down and cools. If this glitch activity has continued over the lifetime of the pulsar, roughly 5% of the moment of inertia of the neutron star must now be effectively removed in the form of pinned vortices; but according to current models⁹, only $\sim 1\%$ of the total moment of inertia is involved in the whole of the superfluid that interpenetrates the inner crust, where pinning might take place.

The alternative explanation, involving a change in the rate of momentum loss through electromagnetic radiation or particle outflow might be accompanied by an observable change in the radiation pattern or the braking index; we intend to search for these in our continued monitoring programme. □

Received 2 June; accepted 24 August 1992.

1. Lyne, A. G. & Pritchard, R. S. *IAU Circ. No.* 4845 (1989).
2. Demianski, M. & Proszynski, M. *Mon. Not. R. astr. Soc.* **202**, 437–451 (1983).
3. Lyne, A. G., Pritchard, R. S. & Smith, F. G. *Mon. Not. R. astr. Soc.* **233**, 667–676 (1988).
4. Boynton, P. E., Groth, E. J., Partridge, R. B. & Williamson, D. T. *IAU Circ. No.* 2179 (1969).
5. Lohsen, E. *Nature* **258**, 688–689 (1975).
6. Lyne, A. G. & Pritchard, R. S. *Mon. Not. R. astr. Soc.* **229**, 223 (1987).
7. Standish, E. M. J. *Astr. Astrophys.* **114**, 297–302 (1982).
8. McCulloch, P. M., Hamilton, P. A., McConnell, D. & King, E. A. *Nature* **346**, 822–824 (1990).
9. Pines, D. & Alpar, M. A. *Nature* **316**, 27–32 (1985).

Curling and closure of graphitic networks under electron-beam irradiation

Daniel Ugarte

Institut de Physique Expérimentale, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

THE discovery¹ of buckminsterfullerene (C_{60}) and its production in macroscopic quantities² has stimulated a great deal of research. More recently, attention has turned towards other curved graphitic networks, such as the giant fullerenes (C_n , $n > 100$)^{3,4} and carbon nanotubes^{5–8}. A general mechanism has been proposed⁹ in which the graphitic sheets bend in an attempt to eliminate the highly energetic dangling bonds present at the edge of the growing structure. Here, I report the response of carbon soot particles and tubular graphitic structures to intense electron-beam irradiation in a high-resolution electron microscope; such conditions resemble a high-temperature regime, permitting a degree of structural fluidity. With increased irradiation, there is a gradual reorganization of the initial material into quasi-spherical particles composed of concentric graphitic shells. This lends weight to the nucleation scheme proposed⁹ for fullerenes, and moreover, suggests that planar graphite may not be the most stable allotrope of carbon in systems of limited size.

The remarkable stability of the C_{60} molecule has been attributed to its highly symmetrical structure where the carbon atoms are arranged at the vertices of a truncated icosahedron¹. The C_{60} molecule may be viewed as a hexagonal graphitic sheet which, by incorporating pentagons, has eliminated all dangling bonds, curling to form a hollow ball (7.1 Å in diameter). For all fullerenes, the strain due to the bending of the sp^2 orbitals tends to concentrate at the vertices of the pentagons. The outstanding stability of the C_{60} molecule is due both to the fact that this is the smallest carbon cage where there are no adjacent pentagons, and to its spherical form which allows the strain to be symmetrically distributed over all atoms¹⁰.

In addition to C_{60} , cylindrical carbon structures have been observed⁵ in which the graphitic sheet has a helical arrangement. The natural question arises: how is it possible to generate such symmetrical, low-entropy forms from the random condensation of carbon vapour? Perhaps it is worthwhile to contemplate the ease with which these carbon hexagonal networks grow as curved or closed sheets rather than the traditionally planar ones. Chemists are conditioned to think of graphitic sheet structures as flat, where carbon atoms are bound in infinite hexagonal sheets, like chicken wire. In carbon vapour, pieces of graphite sheet would have many dangling bonds; they would have little reason to remain flat, and the physical tendency to reach the lowest energy level available would induce the sheets to eliminate their dangling bonds by curling up¹². By heating and properly annealing pure carbon in the absence of other chemically active elements, and under conditions that favour sp^2 carbon network formation, it should be possible to synthesize

curled nets and closed cages. In fact, this is the situation in arc discharges, which is the technique used at present to produce macroscopic quantities of fullerenes², metallofullerenes¹³ and graphitic tubules¹⁴.

Under the conditions of observation in an electron microscope, strong irradiation in some respects resembles a high-temperature regime, allowing structural fluidity; for example, amorphous carbon films usually develop slight graphitization under the electron beam. In particular, irradiation usually heats the sample by energy absorption and ruptures bonds through electron excitations. Furthermore, high-energy particles can transfer momentum to the nuclei, displacing atoms to interstitial lattice sites ('knock on'). Such conditions may be realized by electron bombardment in a high-resolution electron microscope (HREM), and consequently the evolution of a sample may be observed, even up to atomic details under favourable conditions. We must note, however, that electron-beam heating may not lead to the same result as thermal heating, because of the contribution of the excitation processes.

We have irradiated carbon soot in a 300-kV HREM microscope (Philips EM430 ST), using an electron dose up to 10–20 times higher than under normal operating conditions (the usual dose is 10–20 A cm⁻²). Figure 1 shows a sequence of images of a group of irradiated graphitic particles, taken at 10-minute intervals. The original soot, collected in an arc-discharge apparatus, contains mostly nanometric needles formed by coaxial graphitic tubes, with some polyhedral graphitic particles formed by the junction of small flakes of planar graphite (Fig. 1a). All the particles are covered with a thin amorphous carbon

layer. In the intermediate image of the sequence (Fig. 1b), particles now show more marked curvature and, in particular, the tubular structures are collapsing. At this stage, the amorphous carbon layer has graphitized epitaxially onto the particles. Finally (Fig. 1c), the electron annealing leads to a sample composed almost entirely of spherical particles. Detailed examination of the particles shows that they consist of an assembly of concentric spherical graphitic cages (see Fig. 2), the distance between layers agreeing with that for bulk graphite ($d_{002} = 3.34$ Å). The apparent disorder in the spherical shells arises as a consequence of the low electron dose necessary for taking the micrographs; under such conditions, the heating effect of the electron beam is insufficient to permit structural reorganization back to the closed-shell form. Imaging on a much shorter timescale, however, permits the observation of more complete structures (see cover picture). The existence of these structures has already been proposed^{15–17}.

The final spherical structures do not correspond to the deformation of a tube into a sphere with the same number of atoms (see scheme for single-shell particles in Fig. 3a, b), but rather to the formation of a multiple-shell sphere with a very small central cage (Fig. 3c). The reduced dimension of the final inner shell (0.6–1 nm) allows an increase in the number of shells. Following Euler's theorem, any closed hexagonal network contains exactly 12 pentagonal rings; this is the case in the tube (Fig. 3a, the pentagons being situated at the hemispherical domes at the extremities of the cylinder) or in the spherical cage

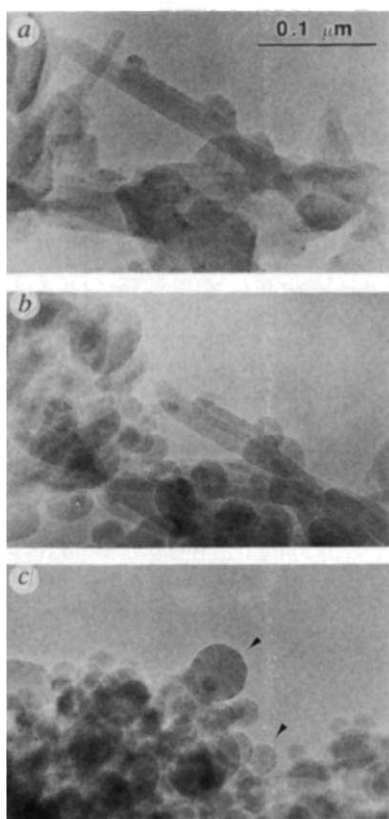


FIG. 1 Sequence of transmission electron micrographs of carbon soot subjected to strong electron beam irradiation. *a*, Original soot containing tubular or polyhedral graphitic particles; *b*, after 10 minutes of strong irradiation, there is a noticeable tendency for the particles forming the soot to become more spherical, especially the graphitic needles; *c*, after 20 minutes, the soot is nearly exclusively composed of quasi-spherical graphitic particles. Further irradiation does not produce significant observable changes in structure.

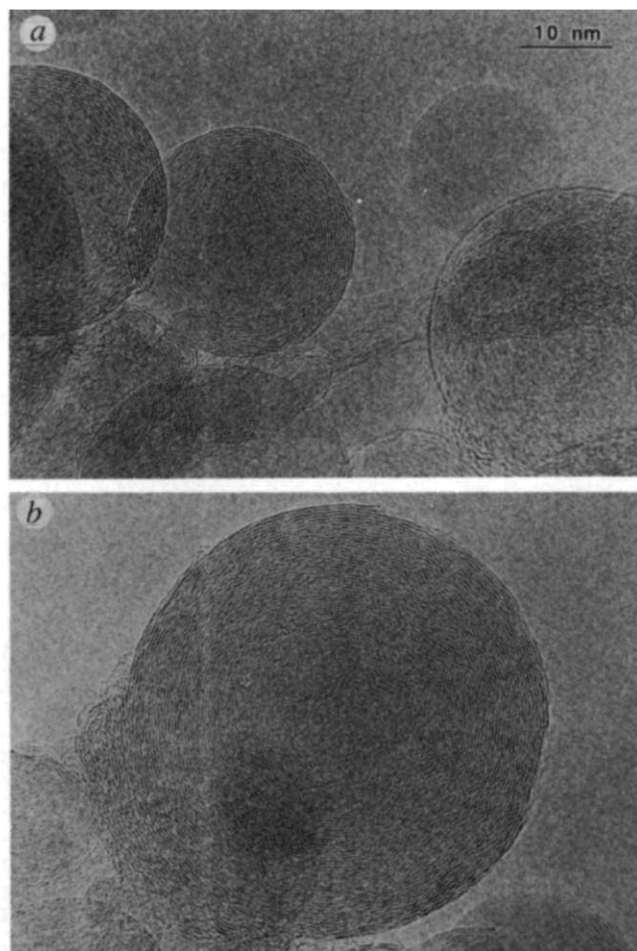


FIG. 2 Detailed structure of the graphitic particles marked with an arrow in Fig. 1c. Dark contrast rings correspond roughly to atomic positions, and the distance between rings corresponds to the (002) lattice parameter of bulk graphite. Note the remarkable sphericity of the particles.

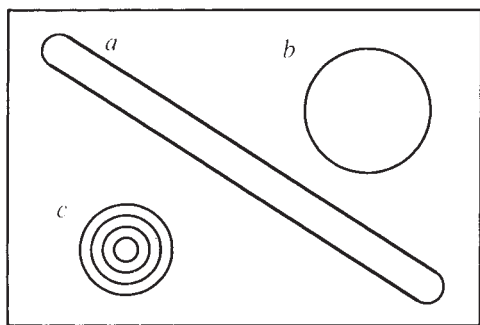


FIG. 3 Schematic representation of three-dimensional particles formed by an equal graphitic surface: *a*, cylindrical structure closed by hemispherical domes (diameter $\varnothing \approx 1$ nm and 13.7 nm long); *b*, spherical cage ($\varnothing \approx 3.74$ nm); *c*, onion-like structure formed by 4 shells ($\varnothing \approx 2.72$ nm), the central one being a C_{60} molecule.

(Fig. 3*b*). The onion-like particle of Fig. 3*c* is formed by four closed shells; in consequence, it contains four times as many pentagons. Our results present clear experimental evidence of the spontaneous tendency of graphite to include pentagons in its hexagonal network and form curved structures. Hence, they support the dangling bond minimization scheme, proposed to explain the growth of fullerenes from carbon vapour⁹.

Curved graphitic sheets can also be formed by irradiation of amorphous carbon particles (Fig. 4). The graphitic structures generated are naturally curled, and two nucleation centres are easily recognizable (marked with arrows in Fig. 4*b*), from which spherical particles will be formed. Further irradiation annealing leads to the separation of the two graphitic spheres.

The formation mechanism for these multiple-shell spheres is based on irradiation-stimulated graphitization, and is rather different from the accretion mechanism originally considered in laser vaporization and arc-discharge experiments, which would produce spiral multiple-shell particles⁹.

The sequence shown in Fig. 1 clearly reveals that if enough energy is provided, spherical structures are favoured over tubular ones. This observation agrees with the predictions made for giant fullerenes^{18,19} (monolayers), but van der Waals interaction between the concentric layers should be included in order to compare calculations with the present experiments.

The 'spherical graphite' that we have observed may attain a

considerable size (47 nm in diameter and ~ 70 shells for the particle shown in Fig. 2*b*). In a few cases, we have even observed spheres several micrometres in diameter, although in this range of sizes a prolonged irradiation period is required before the particles become spherical. We should not rule out the possibility that even larger (possibly macroscopic) graphitic spheres could be generated by adequate annealing of carbon; the maximum size attainable will give us information about the distortion present in the closed graphite sheets. The traditional idea that planar graphite is the most stable form of pure carbon would then have to be seriously reviewed: a flat graphite flake cannot be perfect, and includes many dangling bonds which are usually eliminated by attaching impurities (for example, hydrogen). The 'spherical graphite' presented here is a pure carbon material, which has no dangling bonds, and moreover, having a spherical shape, allows a uniform distribution of the strain because of the out-of-plane geometry. Those of us accustomed to traditional planar graphite, initially surprised by the fascinating fullerenes, are now confronted with supplementary evidence that spherical carbon networks can be favoured under high temperature or strong irradiation regimes.

This notion also raises a point concerning the solid allotropes of carbon. When Krätschmer *et al.*² synthesized large amounts of the C_{60} molecule for the first time, they prepared a new, third form, of solid carbon (called fullerite), which is a three-dimensional packing of C_{60} spheres and is distinct from the two traditional crystalline carbon forms, graphite and diamond. Considering the observed tendency of graphite to form multiple-shell spheres ('onions'), of which single-shell fullerenes are only the first member, we speculate that fullerite is the first member of a family of new solid forms of carbon that could be formed by the packing of these onion-like graphitic spheres, interacting through van der Waals forces. Further experimental work will be needed to produce and isolate multishelled graphitic spheres, but a huge family of carbon materials awaits the skill of experimentalists. □

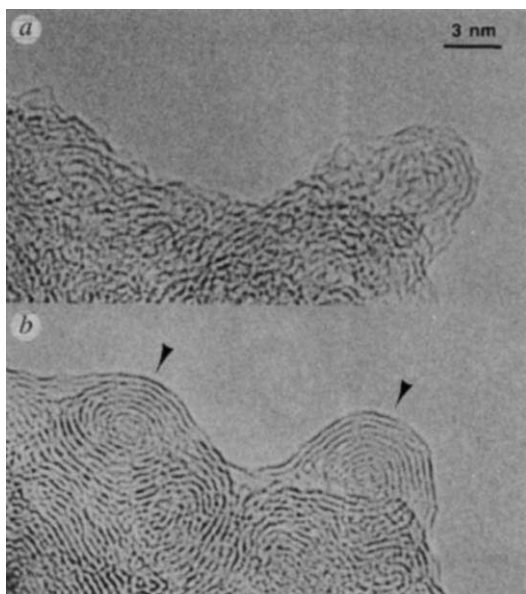


FIG. 4 Transmission electron micrographs of amorphous carbon subjected to electron irradiation. *a*, Original particle; *b*, after 10 minutes of strong irradiation, graphitization (marked with arrows) is present, and the sheets show a clear tendency to form closed cages. Further irradiation would lead to the separation of two graphitic spheres.

Received 7 September; accepted 2 October 1992.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
2. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. *Nature* **347**, 354–358 (1990).
3. Kroto, H. W. & McKay, K. *Nature* **331**, 328–331 (1988).
4. Lamb, L. D. *et al. Science* **255**, 1413–1416 (1992).
5. Iijima, S. *Nature* **354**, 56–58 (1991).
6. Mintimire, J. W., Dunlap, B. I. & White, C. T. *Phys. Rev. Lett.* **68**, 631–634 (1992).
7. Hamada, N., Sawada, S. & Oshiyama, A. *Phys. Rev. Lett.* **68**, 1579–1581 (1992).
8. Tanaka, K., Okahara, K., Okada, M. & Yamada, T. *Chem. Phys. Lett.* **191**, 469–472 (1992).
9. Zhang, Q. L. *et al. J. phys. Chem.* **90**, 525–528 (1986).
10. Kroto, H. W. *Nature* **329**, 529–531 (1987).
11. Bundy, F. P. *Physica A* **156**, 169–178 (1989).
12. Robertson, D. H., Brenner, D. W. & White, C. T. *J. phys. Chem.* **96**, 6133–6135 (1992).
13. Chai, Y. *et al. J. phys. Chem.* **95**, 7564–7568 (1991).
14. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
15. Curl, R. F. & Smalley, R. E. *Scient. Am.* **265**, 32–41 (October, 1991).
16. Iijima, S. *J. Cryst. Growth* **50**, 657–683 (1980).
17. Iijima, S. *J. phys. Chem.* **91**, 3466–3467 (1987).
18. Adams, G. B., Sankey, O. F., Page, J. B., O'Keeffe, M. & Drabold, D. A. *Science* **256**, 1792–1795 (1992).
19. Scuseria, G. E. in *Buckminsterfullerene* (eds Billups, W. E. & Ciufolini, M. A.) (VCH, New York, in the press).

ACKNOWLEDGEMENTS. We thank H. W. Kroto and W. de Heer for discussions and comments, and D. Reinhard and B. D. Hall for critical reading of the manuscript.

Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism

C. T. Kresge*, M. E. Leonowicz*, W. J. Roth*,
J. C. Vartuli* & J. S. Beck†

Mobil Research and Development Corporation,

* Paulsboro Research Laboratory, Paulsboro, New Jersey 08066, USA

† Central Research Laboratory, Princeton, New Jersey 08543, USA

MICROPOROUS and mesoporous inorganic solids (with pore diameters of ≤ 20 Å and ~ 20 –500 Å respectively)¹ have found great utility as catalysts and sorption media because of their large internal surface area. Typical microporous materials are the crystalline framework solids, such as zeolites², but the largest pore dimensions found so far are ~ 10 –12 Å for some metallophosphates^{3–5} and ~ 14 Å for the mineral caxenite⁶. Examples of mesoporous solids include silicas⁷ and modified layered materials^{8–11}, but these are invariably amorphous or paracrystalline, with pores that are irregularly spaced and broadly distributed in size^{8,12}. Pore size can be controlled by intercalation of layered silicates with a surfactant species^{9,13}, but the final product retains, in part, the layered nature of the precursor material. Here we report the synthesis of mesoporous solids from the calcination of aluminosilicate gels in the presence of surfactants. The material^{14,15} possesses regular arrays of uniform channels, the dimensions of which can be tailored (in the range 16 Å to 100 Å or more) through the choice of surfactant, auxiliary chemicals and reaction conditions. We propose that the formation of these materials takes place by means of a liquid-crystal 'templating' mechanism, in which the silicate material forms inorganic walls between ordered surfactant micelles.

Members of this family of materials, designated MCM-41, were first observed in electron micrographs of products from hydrothermal reactions of aluminosilicate gels in the presence of quaternary ammonium surfactants. We prepared the MCM-41 molecular sieve characterized here as follows: 200 g of a solution containing 26 wt% hexadecyltrimethylammonium ion, as $C_{16}H_{33}(CH_3)_3N^+OH/Cl$ ($\sim 30\%$ hydroxide), was combined with 2 g of Catapal alumina, 100 g of tetramethylammonium silicate solution (10% SiO_2 , ratio of tetramethylammonium to $SiO_2 = 1$) and 25 g of a precipitated silica (HiSil), with stirring (molar ratio of $C_{16}H_{33}(CH_3)_3N^+$ to $Si \leq 1$). This mixture was placed in a static autoclave at 150 °C for 48 hours. After cooling it to room temperature, we recovered the solid product by filtration on a Buchner funnel, washed it with water and dried it in air at ambient temperature. The as-synthesized product was then calcined at 540 °C for one hour in flowing nitrogen, followed by six hours in flowing air. The as-synthesized product contains over 40 wt% of the original surfactant as reflected by its composition (molar): 1 N, 19.6 C, 4.7 Si, 0.27 Al. In general, no special precautions (heating rate, atmosphere) are needed during the temperature ramping (from ambient temperature) and calcination processes.

Transmission electron microscopy (Fig. 1a) shows the regular hexagonal array of uniform channels characteristic of MCM-41. A representative electron diffraction pattern (Fig. 1b), with the MCM-41 in the same orientation, confirms the periodicity of the structure. For the sample described above, an interplanar spacing $d_{100} \approx 40$ Å was observed. In the corresponding powder X-ray diffraction pattern of the bulk sample (Fig. 2), the four peaks observed in the low-angle 2θ region can be indexed on a hexagonal unit cell with $a \approx 45$ Å ($2d_{100}/\sqrt{3}$). Generally, both electron and X-ray diffraction patterns show only a few low-order members of the $hk0$ subset of hexagonal reflections. The BET surface area of the preparation is $\approx 1,000$ m² g⁻¹ with exceptionally high sorption capacities of > 50 wt% cyclohexane at

40 torr, 49 wt% *n*-hexane at 40 torr, and 67 wt% benzene at 50 torr. The pore volume of this sample is 0.79 cm³ g⁻¹. The range of pore volumes for MCM-41 samples is 0.7–1.2 cm³ g⁻¹. Figure 3 shows N₂ adsorption isotherms for this material and for an amorphous mesoporous silica. The morphology of MCM-41 depends on synthesis conditions, but it is possible to obtain relatively large (~ 2 μm) hexagonal prisms of MCM-41, as seen in the scanning electron micrograph in Fig. 4.

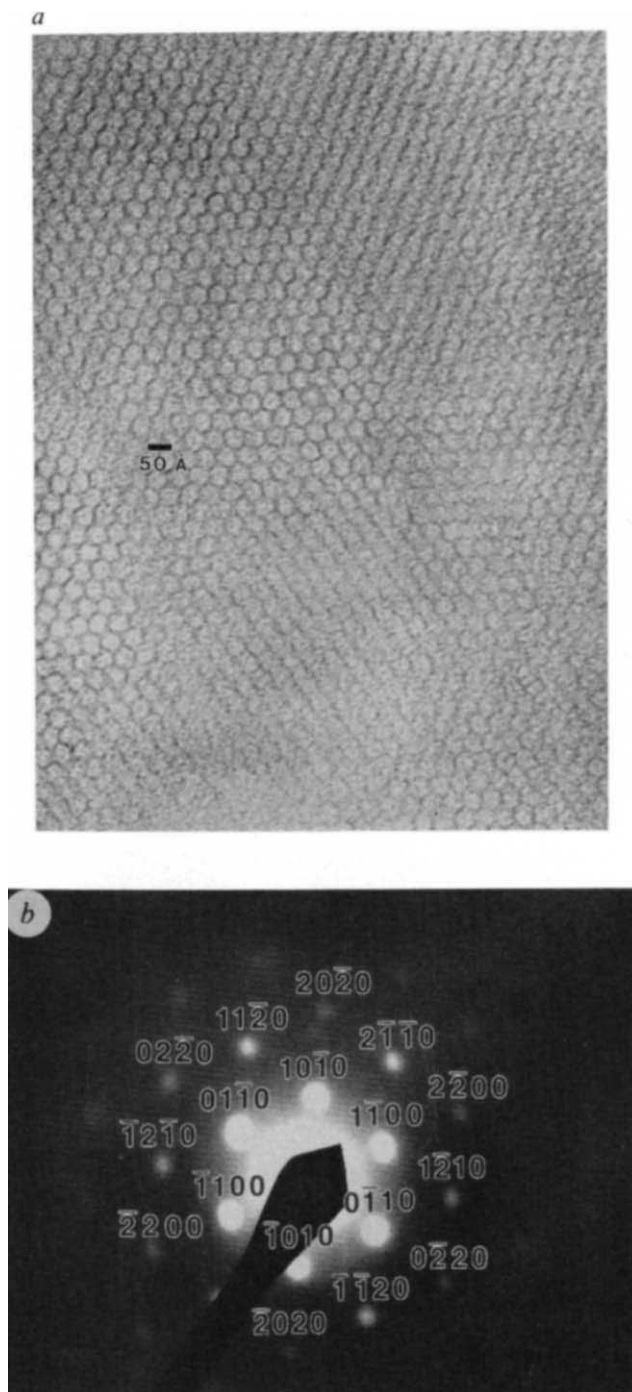


FIG. 1 *a*, Representative transmission electron micrograph of MCM-41. This image was obtained with a JEOL 200CX transmission electron microscope operated at 200 kV from a thin section prepared by ultramicrotomy. The instrument has an interpretable resolution of 4.5 Å and an effective 2-Å objective aperture was used to enhance image contrast. *b*, Representative electron micrograph of MCM-41. Selected area electron diffraction pattern indexed as the $hk0$ projection of a hexagonal unit cell with $a = 45$ Å.

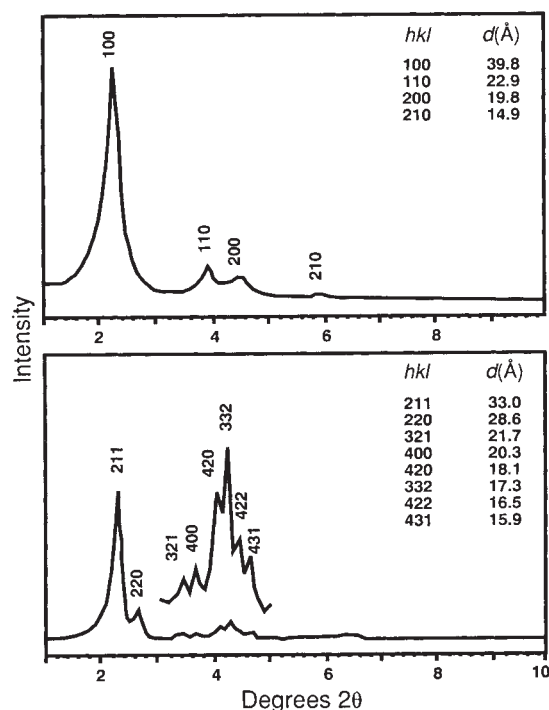


FIG. 2 *a*, Representative X-ray diffraction pattern of MCM-41. The pattern was obtained from a Scintag PAD automated diffraction system using Θ - Θ geometry, Cu $K\alpha$ radiation ($\lambda = 1.5418 \text{ \AA}$) and an energy-dispersive detector. *b*, Representative X-ray diffraction pattern of cubic ($1a3d$) phase. Inset is expansion of region $2\theta = 3-5^\circ$.

The nature of the ordering in the walls of MCM-41—that is, the degree to which the atoms are precisely ordered—is not fully understood. We have not detected any X-ray diffraction peaks with a non-zero l component. The large, regular pore channels of these materials would, however, make the $hk0$ reflections dominate, resulting in little scattering intensity along the c -axis.

The pore diameter of MCM-41 can be varied by changing the alkyl chain length of the cationic surfactants used in the synthesis procedure. For example, by substituting the dodecyltrimethylammonium ion ($\text{C}_{12}\text{H}_{25}(\text{CH}_3)_3\text{N}^+$) for hexadecyltrimethylammonium ion, we produced a sample of MCM-41 with $30\text{-\AA}$ pore diameter. The C/N molar ratio for as-synthesized C_{12} -based MCM-41 is 15, which is consistent

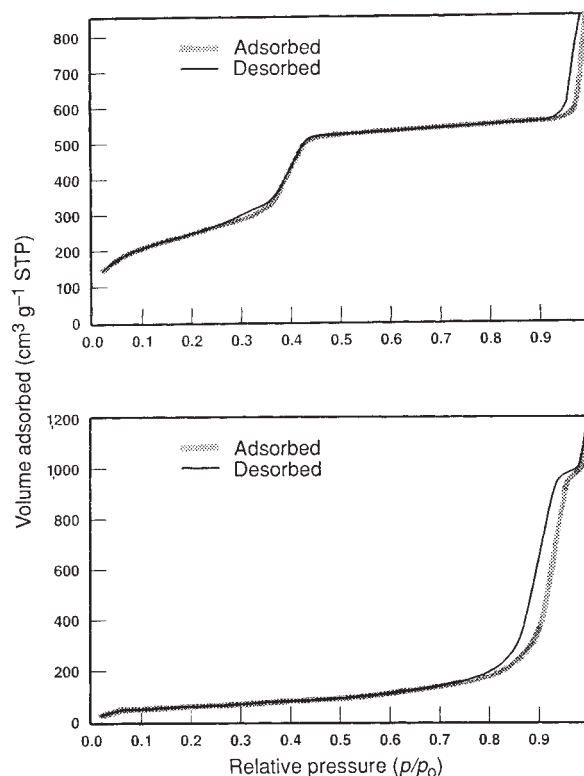


FIG. 3 *a*, N_2 adsorption isotherm for MCM-41. *b*, N_2 adsorption isotherm for amorphous silica (BET surface area $306 \text{ m}^2 \text{ g}^{-1}$). These isotherms were obtained on a Micromeritics Digisorb 2600 adsorption instrument using standard procedures. The isotherm for this MCM-41 shows the inflection characteristic of capillary condensation within the pores, where the p/p_0 (relative N_2 pressure) position of the inflection point is related to the diameter of the pore being filled^{26,27} $\sim 40 \text{ \AA}$. The significant adsorption at lower p/p_0 ($\sim 200 \text{ cm}^3 \text{ g}^{-1}$) is most probably due to monolayer coverage of the walls and not to the presence of microporous phases.

with the surfactant remaining intact. Another way of altering the pore diameter of MCM-41 is to add auxiliary hydrocarbons, such as alkylated benzene (for example 1,3,5-trimethylbenzene), to the synthesis mixture¹⁶. The incremental addition of 1,3,5-trimethylbenzene results in the concomitant increase of d_{100} and the pore diameter. Hexagonal phases with pore diameters up to $\sim 100 \text{ \AA}$ have been characterized.

The microscopy and diffraction results presented above are

FIG. 4 Scanning electron micrograph of a MCM-41 sample. This micrograph was obtained on a JEOL JXA-840 scanning electron microscope using conventional sample preparation and imaging techniques.

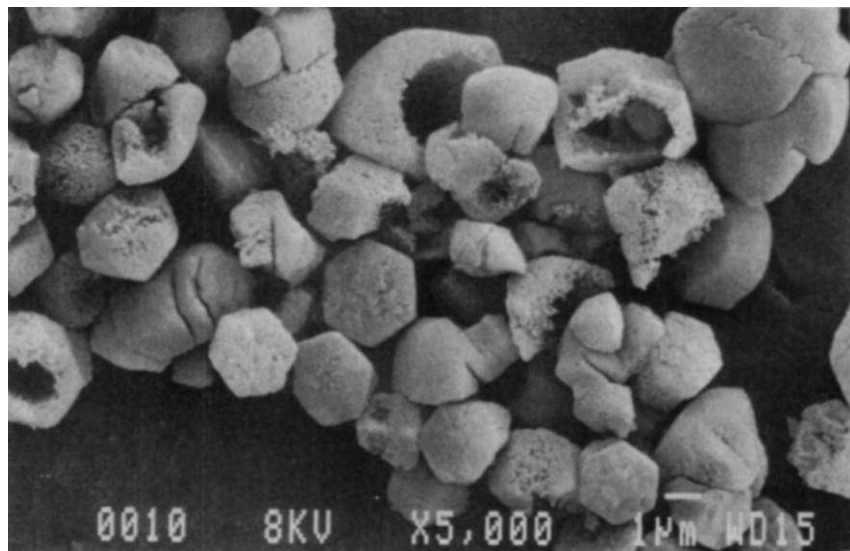
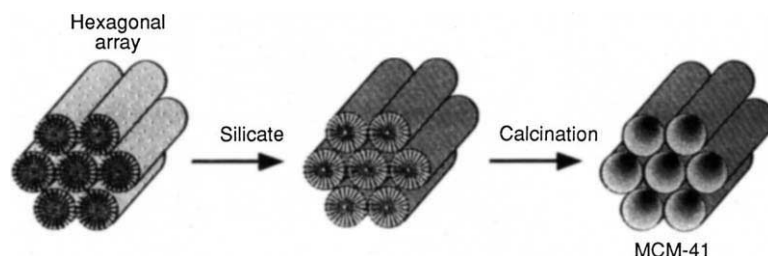


FIG. 5 Schematic drawing of the liquid-crystal templating mechanism. Hexagonal arrays of cylindrical micelles form (possibly mediated by the presence of silicate ions), with the polar groups of the surfactants (light grey) to the outside. Silicate species (dark grey) then occupy the spaces between the cylinders. The final calcination step burns off the original organic material, leaving hollow cylinders of inorganic material.



strikingly similar (R. M. Dessau and C. D. Chang, personal communication) to those obtained¹⁷⁻¹⁹ from lyotropic liquid-crystal phases which are produced in surfactant-water mixtures. These phases are ordered arrays of surfactant aggregates which occur at specific amphiphile concentrations^{20,21}. One such phase, the 'middle' or 'H₁' phase, produces transmission electron microscope images²² analogous to those of MCM-41. The H₁ phase is a hexagonal array of cylindrical micelles in which the hydrophobic hydrocarbon chains are gathered in the centre and the polar groups are arrayed on the surface, in contact with a continuous region of water surrounding the micelles²⁰. The repeat dimensions of the MCM-41 prepared as described above are consistent with those determined for hexadecyltrimethylammonium-based liquid crystals, where cylinder-to-cylinder repeat distances of ~40 Å have been observed²⁰.

The observed dependence on alkyl chain length and the influence of auxiliary organic molecules on the resultant inorganic product are also consistent with two phenomena observed for liquid crystals. The diameter of hexagonal liquid-crystal phases prepared with anionic surfactants depends on the alkyl chain length of the surfactant^{20,21}. Organic species may be solubilized inside the hydrophobic regions of micelles, causing an increase in micelle diameter^{21,23}. Chemicals added to surfactant solutions can increase the porosity of amorphous adsorbents²⁴.

These similarities suggest that these mesoporous molecular sieves are formed by a 'liquid-crystal templating' mechanism. In this mechanism (Fig. 5), inorganic material occupies the continuous solvent (water) region to create inorganic walls between the surfactant cylinders. It may be that encapsulation occurs because anionic aluminosilicate species enter the solvent region to balance the cationic hydrophilic surfaces of the micelles. Alternatively, it may be the introduction of the aluminosilicate species themselves that mediates the hexagonal ordering. Once an ordered array is established, subsequent thermal processing is used to remove the organic material and produce a stable mesoporous molecular sieve.

Although other workers¹³ have produced mesoporous silicates with high surface areas, the mechanism used has been the intercalation of a guest (such as hexadecyltrimethylammonium chloride) into a layered host (such as kanemite). The mesopores of these materials are not an ordered regular array (as observed in MCM-41). The final product retains, in part, the layered nature of the kanemite precursor. Furthermore, the concentrations, reagents, pH and reaction conditions are not conducive for liquid-crystal formation and templating.

Additional strong evidence for the proposed liquid-crystal templating mechanism is the discovery of another member of this family^{14,15} that mimics a liquid-crystal structure²⁵. A representative powder X-ray diffraction pattern of a silicate molecular sieve with cubic symmetry (*Im3d*) is given in Fig. 2. The cubic phase is prepared by increasing the molar ratio of C₁₆H₃₃(CH₃)₃N⁺ to Si to values greater than one. □

- Estermann, M., McCusker, L. B., Baerlocher, C., Merrouche, A. & Kessler, H. *Nature* **352**, 320-323 (1991).
- Moore, P. B. & Shen, J. *Nature* **306**, 356-358 (1983).
- Iler, R. K. *The Chemistry of Silica* (Wiley, New York, 1979).
- Pinnavaia, T. J. *Science* **220**, 365-371 (1983).
- Landis, M. E. *et al.* *J. Am. chem. Soc.* **113**, 3189-3190, 1991.
- Vaughan, D. E. W. & Lussier, R. J. *Proc. 5th Int. Conf. Zeolites* (ed. Rees, L. V. C.) 94-100 (Heyden, London, 1980).
- Vaughan, D. E. W. *Am. chem. Soc. Symp. Series* **368**, 308-325 (1988).
- Tindwa, R. M., Ellis, D. K., Peng, G. Z. & Clearfield, A. *J. chem. Soc. Faraday Trans. 1*, **81**, 545-548 (1985).
- Yanagisawa, T., Shimizu, T., Kazuyuki, K. & Kato, C. *Bull. Chem. Soc. Jpn* **63**, 988-992 (1990).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J. & Vartuli, J. C. *U.S. Patent No. 5,098,684* (1992); *U.S. Patent No. 5,102,643* (1992).
- Beck, J. S. *et al.* *U.S. Patent No. 5,108,725* (1992).
- Beck, J. S. *U.S. Patent No. 5,057,296* (1991).
- Ekwall, P. *Advances in Liquid Crystals*, Vol. 1 (ed. Brown, G. H.) (Academic, New York, 1971).
- Ekwall, P., Mandell, L. & Fontell, K. *Liquid Crystals* (ed. Brown, G. H.) 325-334 (Gordon and Breach, London, 1969).
- Luzzati, V. *Biological Membranes* (ed. Chapman, D.) 71-123 (Academic, New York, 1968).
- Tiddy, G. J. T. *Phys. Rep.* **57**, No. 1, 1-46 (1980).
- Winsor, P. A. *Chem. Rev.* **68**, No. 1, 1-40 (1968).
- Goodman, J. F. & Clunie, J. S. *Electron Microscopy of Liquid Crystals, Liquid Crystals and Plastic Crystals*, Vol. 2 (eds. Gray, G. W. & Winsor, P. A.) 1-23 (Wiley, New York, 1974).
- Speight, P. P. A., Skoulios, A. E. & Luzzati, V. *Acta cryst.* **14**, 866-872 (1961).
- Komarov, V. S. & Kuznetsova, T. F. *Vesti Akad. Nauk BSSR* No. 2, 22-27 (1978).
- Luzzati, V. & Speight, P. P. A. *Nature* **215**, 701-704, 1967.
- Gregg, S. J. & Sing, K. S. W. *Adsorption, Surface Area, and Porosity*, 2nd Edn (Academic, New York, 1982).
- Barrett, E. P., Joyner, L. G. & Halenda, P. P. *J. Am. chem. Soc.* **73**, 373-380 (1951).

ACKNOWLEDGEMENTS. We thank the technical staff of the Paulsboro and Central Research Laboratories, Mobil Research and Development Corporation, for discussions. In particular, we thank C. D. Chang, R. M. Dessau and H. M. Princen for alerting us to references and for discussions, D. T. Geston and K. G. Simmons for assistance, J. B. Higgins and N. H. Goeke for the scanning electron micrograph, J. L. Schlenker for assistance in X-ray diffraction pattern indexing and Mobil Research and Development Corporation for its support.

Anthropogenic influence on the distribution of tropospheric sulphate aerosol

J. Langner*, H. Rodhe†, P. J. Crutzen‡
& P. Zimmermann‡

* Swedish Meteorological and Hydrological Institute, S-601 76 Norrköping, Sweden

† Department of Meteorology, Stockholm University, S-106 91 Stockholm, Sweden

‡ Max-Planck Institute for Chemistry, P.O. Box 3060, D-6500 Mainz, Germany

HUMAN activities have increased global emissions of sulphur gases by about a factor of three during the past century, leading to increased sulphate aerosol concentrations, mainly in the Northern Hemisphere. Sulphate aerosols can affect the climate directly, by increasing the backscattering of solar radiation in cloud-free air, and indirectly, by providing additional cloud condensation nuclei¹⁻⁴. Here we use a global transport-chemistry model to estimate the changes in the distribution of tropospheric sulphate aerosol and deposition of non-seasalt sulphur that have occurred since pre-industrial times. The increase in sulphate aerosol concentration is small over the Southern Hemisphere oceans, but reaches a factor of 100 over northern Europe in winter. Our calculations indicate, however, that at most 6% of the anthropogenic sulphur emissions is available for the formation of

Received 1 May; accepted 1 September 1992.

- IUPAC Manual of Symbols and Terminology *Pure appl. Chem.* **31**, 578 (1972).
- Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types*, 2nd Edn (Butterworths, London, 1988).
- Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. *Nature* **331**, 698-699 (1988).
- Dessau, R. M., Schlenker, J. L. & Higgins, J. B. *Zeolites* **10**, 522-524 (1990).

new aerosol particles. This is because about one-half of the sulphur dioxide is deposited on the Earth's surface, and most of the remainder is oxidized in cloud droplets so that the sulphate becomes associated with pre-existing particles. Even so, the rate of formation of new sulphate particles may have doubled since pre-industrial times.

We use the three-dimensional Eulerian transport model MOGUNTIA⁵ in our calculations. This is based on monthly mean meteorological data and covers the global troposphere with a horizontal resolution of 10° longitude by 10° latitude, and with 10 vertical layers between the surface and 100 hPa. The present version simulates the tropospheric distribution of dimethyl sulphide (DMS), sulphur dioxide (SO₂) and non-sea-salt (n.s.s.) sulphate aerosol (SO₄²⁻) (ref. 6). Most volatile sulphur compounds are emitted as DMS and SO₂. The model assumes that SO₄²⁻ is the end product.

We have made calculations for industrial and pre-industrial sulphur emissions. The assumed distributions are shown in Fig. 1 (for details see ref. 6). Recent estimates of the global industrial input of sulphur for the year 1980 are within ±15% of the value used in ref. 6. The uncertainty in the estimate of the global natural emission of sulphur gases is about a factor of two, mainly because of uncertainties in the emission of DMS from the oceans. For this flux we have chosen a value of 25 Tg S yr⁻¹, in the middle of the range suggested by recent estimates^{7,8}. The distribution of the natural emissions and their variation with time are uncertain because of incomplete information about oceanic DMS emissions and variable volcanic emissions (see Figs 2 and 3 for details).

Gas-phase oxidation of DMS to SO₂ and of SO₂ to sulphate is proportional to the concentrations of hydroxyl radicals (OH). Three-dimensional time-dependent OH fields for both industrial and pre-industrial situations are derived from a version of the MOGUNTIA model that contains full CH₄-CO-NO_x-OH-O₃ chemistry⁹ and takes into account changes in CH₄, CO and NO.

We assume that aqueous-phase oxidation of SO₂ to sulphate is proportional to cloudiness, wet deposition of SO₂ and sulphate is proportional to precipitation, and dry deposition of SO₂ and sulphate is proportional to their surface concentrations. Further details are given in ref. 6.

Figure 2 shows distributions of sulphate in the lowest layer (0 to 500 m) in January and July for the industrial and pre-industrial scenarios. The impact of anthropogenic emissions is evident over the Northern Hemisphere. The main source regions in Europe, North America and China show up with maximum values exceeding 1 p.p.b.v. (10⁻⁹ mol sulphate per mol air). The average atmospheric lifetime of ~5 days for SO₄²⁻ is long enough to allow for a considerable transport out of the source regions. This can be seen most clearly east of North America and China and to the east and southeast of Europe. Anthropogenic influence is also evident in the Southern Hemisphere over Africa, South America and Australia.

When anthropogenic emissions are excluded, the highest sulphate concentrations occur in tropical regions where emissions of DMS are higher. The influence of volcanic emissions is marked over Indonesia and over the western part of the North Pacific. Minimum values below 10 p.p.t.v. (10⁻¹² mol sulphate per mol air) are calculated in winter over polar areas and over central and northern Asia. According to the calculations, the tropospheric n.s.s.-sulphate burden has increased from 190 to 530 Gg S (1 Gg = 10⁹ g) in the Northern Hemisphere since the start of industrial times. The corresponding change in the Southern Hemisphere is 160–230 Gg S.

Because sulphate aerosols may affect the reflectivity of low clouds (mainly stratus), we now estimate the impact of anthropogenic emissions on the sulphate concentration at the level of such clouds (500–1,500 m). Figure 3 shows the ratio of the sulphate concentrations for the industrial and pre-industrial eras for July and January. The present sulphate concentrations at the level of low clouds are at least a factor of two larger than

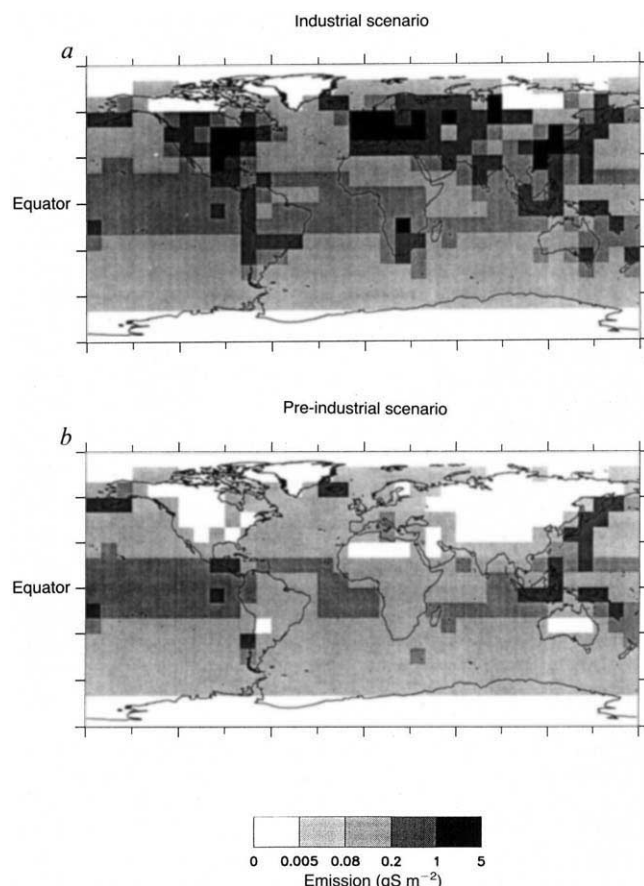


FIG. 1. Distribution of annual sulphur emissions used in the model simulations for (a) the industrial and (b) the pre-industrial era. Total global fluxes (Tg S yr⁻¹) for the industrial scenario are (Southern Hemisphere fluxes in parenthesis): industrial (95% SO₂, 5% SO₄²⁻), 70 (6); biomass burning (SO₂), 2.5 (1.4); oceans (DMS), 25 (14.2); volcanoes (SO₂), 8.5 (2.7); soil and plants (DMS, H₂S), 1 (0.4). For pre-industrial times we reduced the biomass burning by 90%, roughly reflecting the change in tropical population since the beginning of the last century.

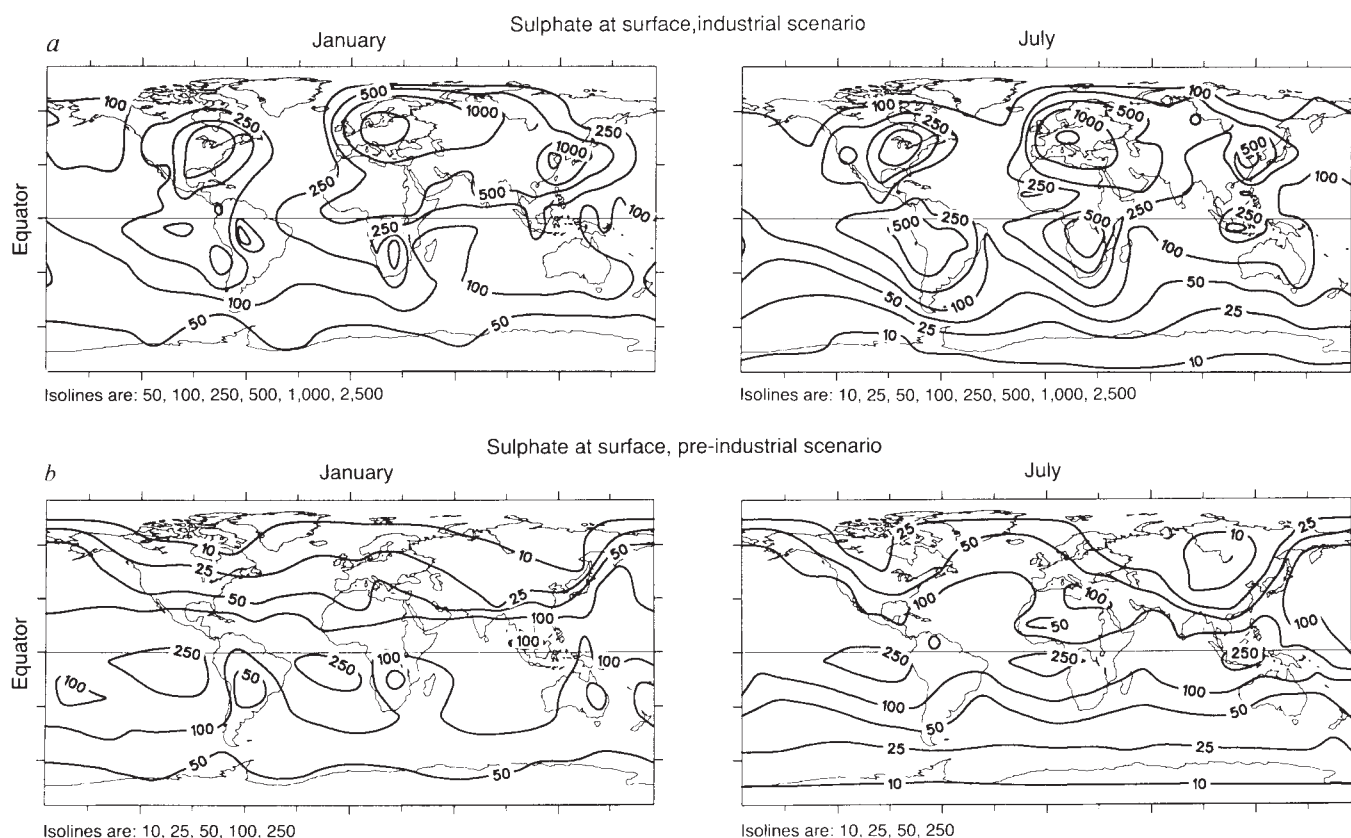


FIG. 2 Calculated distribution of SO_4^{2-} in the lowest 500 m for (a) the industrial and (b) the pre-industrial eras for January (left) and July (right). Units: p.p.t.v. Calculated values for the industrial scenario are within $\pm 50\%$ of observed values in Europe and North America⁶. Uncertainties related to the wet deposition parameterization range from $\pm 20\%$ over industrialized

areas to $\pm 50\%$ over remote regions⁶. The total uncertainty outside industrialized regions is larger, mainly because of uncertainties in natural emissions. In general the calculated concentrations in remote regions and for the pre-industrial scenario are estimated to have a factor of two uncertainty although locally the uncertainty may be larger.

in the pre-industrial era, in July (January), over 7% (27%) of the Pacific and over 35% (74%) of the Atlantic. The corresponding numbers for the Northern Hemisphere as a whole are 51% (68%). Increases by two orders of magnitude are calculated over the most polluted parts of Europe. We calculate a decrease in cloud-level sulphate concentrations over remote ocean areas, because of the reduction in OH concentrations in these regions caused by higher concentrations of CH_4 and CO . More SO_2 is lost through surface deposition, and thus less SO_4^{2-} is produced. The decrease is generally 10–20% but covers a large area.

The calculated distributions of the annual deposition of sulphur (wet plus dry) for the industrial and pre-industrial scenarios (not illustrated) are similar to those of the concentration fields with fluxes exceeding $2.5 \text{ g S m}^{-2} \text{ yr}^{-1}$ in areas heavily affected by anthropogenic emissions. Figure 3c shows the ratio of the total n.s.s.-sulphur deposition between the industrial and the pre-industrial scenario. The total deposition of sulphur is calculated to have increased by a factor of two or more over 58% of the Northern Hemisphere and over 16% of the Southern Hemisphere. Increases by almost two orders of magnitude are calculated in the regions most heavily affected by anthropogenic emissions in the Northern Hemisphere. Large changes can also be seen in the main anthropogenic source regions in the Southern Hemisphere.

Little information is available to verify the calculated changes in the tropospheric distribution of SO_4^{2-} and fluxes of n.s.s. sulphur presented here. This is especially true regarding concentrations of n.s.s. sulphate in air. A survey of changes in deposition fluxes of sulphur is given by Brimblecombe *et al.*¹⁰ Data on sulphate in river runoff, soil, peatbog and lake sediments

all qualitatively indicate significant changes since pre-industrial times in the deposition of sulphur to the continents in the Northern Hemisphere. The best quantitative information comes from ice-core records obtained in Greenland showing an enhancement by a factor of 2–4 over the last hundred years¹¹. This agrees well with the calculated change by ~ 2 –5 in total deposition of n.s.s. sulphur in this region.

Changes in the concentration of sulphate aerosol may affect climate. From empirical relations between light extinction and sulphate concentration, the direct backscattering effect of the anthropogenic increase in sulphate burden has been estimated to cause a reduction of incoming solar radiation at the surface by $\sim 1 \text{ W m}^{-2}$ averaged over the Northern Hemisphere¹.

The indirect effect through cloud albedo changes is more difficult to quantify, as it is related to changes in the number concentration of cloud condensation nuclei (CCN) rather than to changes in the mass concentration of aerosol sulphate. Because the sulphate produced through aqueous-phase oxidation becomes associated with pre-existing particles, this process cannot produce new particles, although it can modify the size distribution of the aerosol so that additional particles may be activated at a given supersaturation¹². The part of the emitted SO_2 that can lead to new particles is therefore the part that is oxidized by OH radicals in the gas phase. This part is calculated to have increased by 4.2 (from 4.8 to 9) Tg S yr^{-1} since pre-industrial times. Only $\sim 6\%$, of the $72.2 \text{ Tg S yr}^{-1}$ emitted from anthropogenic activities is therefore available for production of new particles. The corresponding figure for the natural emissions is 14%. This low efficiency is because about one-half of the SO_2 is deposited on the Earth's surface and most of the remainder

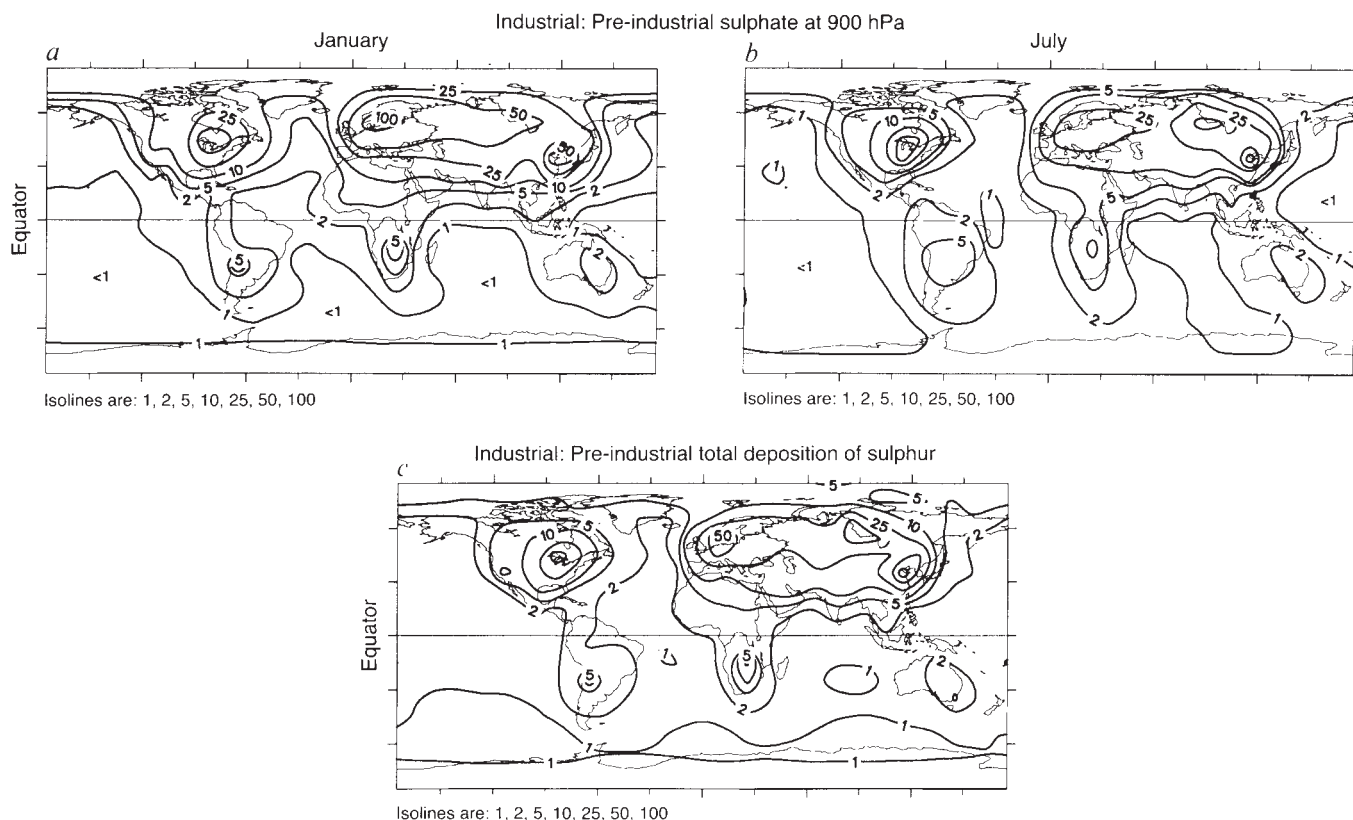


FIG. 3 Ratio of the SO_4^{2-} concentration at about 1 km (900 hPa) for the industrial scenario to that for the pre-industrial scenario, in (a) January and (b) July. c, Ratio of the total (wet plus dry) n.s.s.-sulphur deposition for the industrial to the pre-industrial scenario. Uncertainties related to the wet deposition parameterization tend to cancel when calculating ratios and are small in areas dominated by either natural or anthropogenic sources, and

$\pm 20\%$ in areas where both natural and anthropogenic sources are important. The total uncertainty is larger, mainly because of uncertainties in natural emissions. If we assume a factor of two uncertainty in natural emissions, we can calculate the following ranges (in parenthesis) for each isoline: 2 (1.5–3), 5 (3–9), 10 (5.5–19), 25 (13–49).

is oxidized in cloud droplets. It is possible that the OH oxidation is even less important than we have estimated, as oxidation of SO_2 may also take place in seasalt particles¹³ and soil-derived particles. On the other hand, a small fraction of the industrial emission is likely to occur as sulphate. The OH oxidation varies substantially in space and time. It is largest in the boundary layer and in the upper troposphere during summer, whereas aqueous-phase oxidation dominates during winter in the lower and middle troposphere where clouds are abundant. This information can only be taken as an indication as to where new particle production from SO_2 oxidation may be important. Detailed calculations of aerosol dynamics are needed to establish changes in number concentrations.

Estimates of the possible climate impact of anthropogenically derived CCN have been based on a variety of assumptions. Kaufmann *et al.*¹⁴ assumed, in contradiction to our results, that 40% of the fossil-fuel SO_2 emissions resulted in new particles. They further assumed that about one-half of these particles are active as CCN and concluded that the indirect effect would probably dominate over the greenhouse effect from fossil-fuel CO_2 . Charlson *et al.*³ and Schwartz¹⁵ assumed that the CCN numbers in the Northern Hemisphere were 30–50% higher than in the Southern Hemisphere. A 30% increase in the number of CCN could imply a negative climate forcing at least as large as the direct sulphate forcing¹⁵, resulting in a total negative sulphate aerosol forcing of the same magnitude as the positive forcing due to increases in CO_2 (1.5 W m^{-2} ; ref. 16) and other greenhouse gases (1.0 W m^{-2}). Such an assumption is not contradicted by our estimates but it may still be a substantial overestimate. In fact, Wigley and Raper¹⁷ argue that such a large total aerosol

forcing is incompatible with observed changes in hemispheric mean temperatures. We note that the difference in forcing between the hemispheres could have been augmented by the decrease in sulphate production over much of the Southern Hemisphere because of reductions in OH concentrations, as discussed above.

Although the possible climate effects of increased sulphur emissions certainly deserve attention, we should also keep in mind the acidifying effects of the sulphur deposition. Sulphur emissions have been reduced in Europe and North America, but sulphur deposition still exceeds the critical levels that lead to acidification over much of the Northern Hemisphere. Using $0.5 \text{ g S m}^{-2} \text{ yr}^{-1}$ as the critical load for sulphur deposition¹⁸, we find that this value is exceeded over 20% of the Northern Hemisphere continents. □

Received 13 May; accepted 2 September 1992.

1. Charlson, R. J., Langner, J. & Rodhe, H. *Tellus A/B* **43**, 152–163 (1991).
2. Twomey, S. *Atmospheric Aerosols* (Elsevier, Amsterdam, 1977).
3. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
4. Wigley, T. M. L. *Nature* **339**, 365–367 (1989).
5. Zimmermann, P. H. in *Proc. 16th NATO/CCMS Int. Tech. Mtg Air Pollution Modeling and its Application*, 593–608 (Reidel, Dordrecht, 1987).
6. Langner, J. & Rodhe, H. *J. Atmos. Chem.* **13**, 225–263 (1991).
7. Andreae, M. O. *Mar. Chem.* **30**, 1–29 (1990).
8. Bates, T. S., Cline, J. D., Gammon, R. H. & Kelly-Hansen, S. R. *J. geophys. Res.* **92**, 2930–2938 (1987).
9. Crutzen, P. J. & Zimmermann, P. H. *Tellus A/B* **43**, 136–151 (1991).
10. Brimblecombe, P., Hammer, C., Rodhe, H., Ryaboshapko, A. & Bontou, C. F. in *Evolution of the Global Biogeochemical Sulphur Cycle* (eds Brimblecombe, P. & Lein, A. Y.) 77–121 (Wiley, Chichester, 1989).
11. Neftel, A., Beer, J., Oeschger, H., Zürcher, F. & Finkel, R. C. *Nature* **313**, 611–613 (1985).
12. Hegg, D. A. *Geophys. Res. Lett.* **17**, 21–65 (1990).
13. Sievering, H. *et al. Atmos. Environ.* **A25**, 1479–1487 (1991).
14. Kaufmann, Y. J., Fraser, R. S. & Mahoney, R. L. *J. Clim.* **4**, 578–588 (1991).
15. Schwartz, S. E. *Nature* **336**, 441–445 (1988).

16. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change: The IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1990).
 17. Wigley, T. M. L. & Raper, S. C. B. *Nature* **357**, 293–300 (1992).
 18. Henriksen, A. & Brakke, P. F. *Envir. Sci. Technol.* **22**, 8–14 (1988).

ACKNOWLEDGEMENTS. This research is part of the EUROTRAC subproject GLOMAC and is sponsored by the Swedish National Environment Protection Board and the German Klimaforschungsprogramm. Comments by F. Raes improved the manuscript.

Effects of boreal forest vegetation on global climate

Gordon B. Bonan*, David Pollard
& Starley L. Thompson

National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307–3000, USA

TERRESTRIAL ecosystems are thought to play an important role in determining regional and global climate^{1–6}; one example of this is in Amazonia, where destruction of the tropical rainforest leads to warmer and drier conditions^{4–6}. Boreal forest ecosystems may also affect climate. As temperatures rise, the amount of continental and oceanic snow and ice is reduced, so the land and ocean surfaces absorb greater amounts of solar radiation, reinforcing the warming in a 'snow/ice/albedo' feedback which results in large climate sensitivity to radiative forcings^{7–9}. This sensitivity is moderated, however, by the presence of trees in northern latitudes, which mask the high reflectance of snow^{10,11}, leading to warmer winter temperatures than if trees were not present^{12–14}. Here we present results from a global climate model which show that the boreal forest warms both winter and summer air temperatures, relative to simulations in which the forest is replaced with bare ground or tundra vegetation. Our results suggest that future redistributions of boreal forest and tundra vegetation (due, for example, to extensive logging, or the influence of global warming) could initiate important climate feedbacks, which could also extend to lower latitudes.

The global climate model¹⁵ combines atmospheric general circulation with transfer of energy, moisture and momentum between the atmosphere and oceans and land surfaces. The atmospheric general circulation model is derived from the National Center for Atmospheric Research community climate model CCM1¹⁶, but with new treatments of clouds¹⁷, penetrative plume convection, planetary boundary-layer mixing, solar radiation and its diurnal cycle¹⁸, and semi-Lagrangian transport of water vapour^{19–21}. The surface models include vegetation transfer schemes similar to the biosphere-atmosphere transfer scheme BATS²² and the simple biosphere model SiB²³, multilayer models of heat and moisture flow in soil, snow and sea ice, and a slab ocean mixed layer²⁴. Important features of the vegetation model are: vegetation is represented by upper and lower canopy layers; the absorption, transmission and reflection of solar radiation in each layer are calculated from the two-stream approximation, with separate calculations for direct and diffuse radiation in visible and near-infrared wavelengths; wind is modelled using mixing-length logarithmic profiles above and between layers and a simple diffusive model within each layer; water and snow are intercepted by and cascade through the canopy; and irradiance, temperature, vapour pressure and soil moisture control stomatal resistance. The soil model simulates soil moisture, temperature and permafrost. The snow model simulates varying snow depth and fractional areal cover on the ground. The atmospheric general circulation model has a horizontal resolution of $\sim 4.5^\circ$ latitude by 7.5° longitude, whereas all surface models use a horizontal grid of 2° by 2° . Soil properties are

assigned on the basis of their texture and colour as in BATS²², and vegetation attributes are prescribed as in SiB²³.

The boreal forest is a broad circumpolar mixture of evergreen needleleaf, deciduous needleleaf and deciduous broadleaf tree species, occupying $11.6 \times 10^6 \text{ km}^2$ in Arctic and sub-Arctic regions²⁶. We did simulations with the boreal forest present (as a control) and with all forests north of 45° N replaced by bare ground, which is equivalent to moving the treeline between boreal forest and tundra vegetation southwards. We used the same soil properties in both sets of simulations. Although this perturbation is extreme, it is an appropriate way of isolating the climatic significance of the boreal forest. Moreover, extensive logging operations may be common in the future, especially in Siberia²⁷.

With interactive oceans, the removal of the forest vegetation increased the land surface albedo during January and April (Fig. 1a, b), causing colder air temperatures than in the control simulation (Fig. 2a, b). The largest change in air temperature occurred in April, when land surfaces were as much as 12° C colder (Fig. 2b). The air continued to be as much as 5° C colder in July (Fig. 2c), despite little change in albedo (Fig. 1c). October was 5° C colder at 60° N (Fig. 2d), and albedo increased as snow cover increased (Fig. 1d).

The changes in temperature of the upper snow and soil layers followed the same pattern as air temperature. The moisture content of the upper soil layer (0 to 5 cm depth) was little changed from the control simulation for January and July. The soil was drier between 45 and 55° N in April, when snow began to melt in the control simulation but not in the deforested simulation. October was also drier from 55 to 65° N , where colder air temperatures in the deforested simulation caused precipitation to fall as snow rather than as rain as in the control simulation. Similar patterns of moisture change occurred in the next two soil layers (5–15 cm depth and 15–35 cm depth).

The perturbation caused by replacing existing forest with bare ground is the largest possible change we could make. We examined the sensitivity of our results to smaller perturbations through additional experiments in which (1) the boreal forest was replaced with tundra vegetation and (2) the soil in the bare-ground simulation was made as dark as possible. Neither of these significantly altered the changes in air temperatures caused by deforestation. In particular, the short stature, low leaf

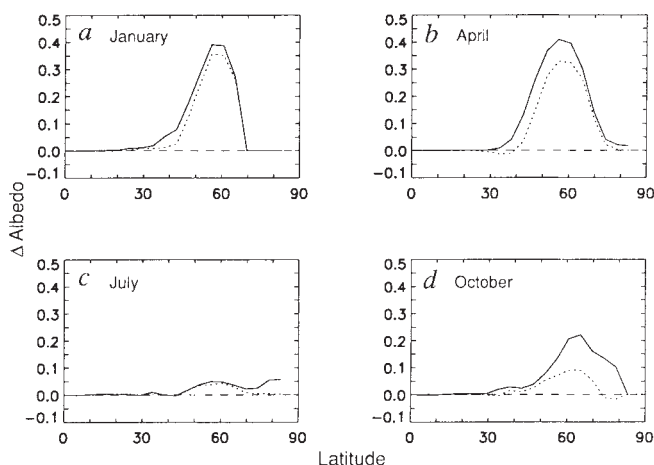


FIG. 1 Zonally averaged surface albedo over land. Albedos are expressed as the difference between the deforested and control simulations. a, January; b, April; c, July; d, October. (—), Simulations with interactive SST and sea ice. Control and deforested simulations were each integrated for 15 years, and albedos were averaged over the last five years. (---), Simulations with prescribed SST and sea ice. Control and deforested simulations were each integrated for five years, and albedos were averaged over the last four years.

* To whom correspondence should be addressed.

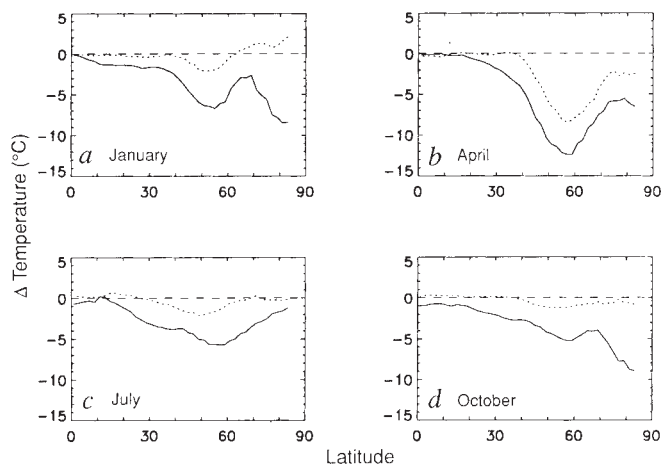


FIG. 2 Zonally averaged air temperature at a height of 2 m above land surfaces. See Fig. 1 for details.

area index and low fractional cover of tundra vegetation prevent it masking the snow albedo.

The summer climate cooling was caused in part by oceanic influences. The colder winter air temperatures caused by deforestation reduced the sea surface temperatures (SST) in Arctic regions. The oceans introduced a thermal lag which inhibited warming in the summer. In addition, the extent of sea ice increased because of the colder winter climate, reinforcing the cooling because of higher ocean albedos.

To examine the importance of the oceans, we did an additional experiment in which SST and sea ice were prescribed from ref. 28 as in the standard CCM1¹⁶. With prescribed SST and sea ice, the climate cooling caused by the deforestation was much less than that simulated with interactive oceans (Fig. 2). In particular, roughly one-third to one-half of the July cooling was due to interactive oceans (Fig. 2c). This implies that the winter-spring cooling pulse caused by boreal deforestation is perpetuated year-round by the thermal reservoir of the oceans and is amplified by the positive feedback due to sea ice.

The effects of the boreal forest on climate extended well beyond the region where the forest was removed. Air temperatures at 30° N were 1.6 to 3.2 °C cooler throughout the year (Fig. 2). Even at 10° N, air temperatures were 1 °C cooler in January and October (Fig. 2). These hemispheric climate changes are much larger than those due to deforestation in the Amazon⁴⁻⁶.

Sensitivity analyses with the land surface model alone (without reciprocal coupling to the climate model) indicated that deforestation should instead create a warmer land surface climate on calm summer days. The different predictions given by the land surface model used alone, used with the fully coupled climate model, and used with prescribed SST and sea ice show that the effect of deforestation on summer air temperatures cannot be resolved by studying only one component process of the system. The climate of northern latitudes is a highly interactive system in which coupling among the atmosphere, oceans and land surfaces cannot be ignored.

Research on biosphere-atmosphere interactions in Arctic and sub-Arctic regions has mainly focused on biogeochemical feedbacks: boreal forest ecosystems contain large quantities of organic carbon^{26,29}, affect the seasonal dynamics of atmospheric CO₂ (refs 30, 31) and can be large annual carbon sinks³². The sensitivity of high-latitude climate to increasing atmospheric concentrations of CO₂ (ref. 33) has focused attention on carbon feedbacks to the atmosphere that may accentuate climate change^{29,33,34}. Our studies suggest that the physical effects of boreal forests on climate are as significant as the expected biogeochemical effects.

Historically, it has been thought that the extent and nature of the boreal forest were determined by climate³⁵⁻³⁹. Contours of climate indices do in fact outline the geographical boundaries of the boreal forest and its subdivisions³⁵⁻³⁷. But our results and previous studies¹²⁻¹⁴ indicate that the locations of the boreal forest and of correlated climate indices are the outcome of coupled dynamical interactions in which the geographical distribution of the boreal forest affects climate and vice versa. For example, the current northern and southern boundaries of the boreal forest are correlated with the positions of the July 13 °C and July 18 °C isotherms, respectively³⁶. Assuming that the July isotherms are causal, the summer cooling caused by deforestation is sufficient to prevent forest regrowth in much of the

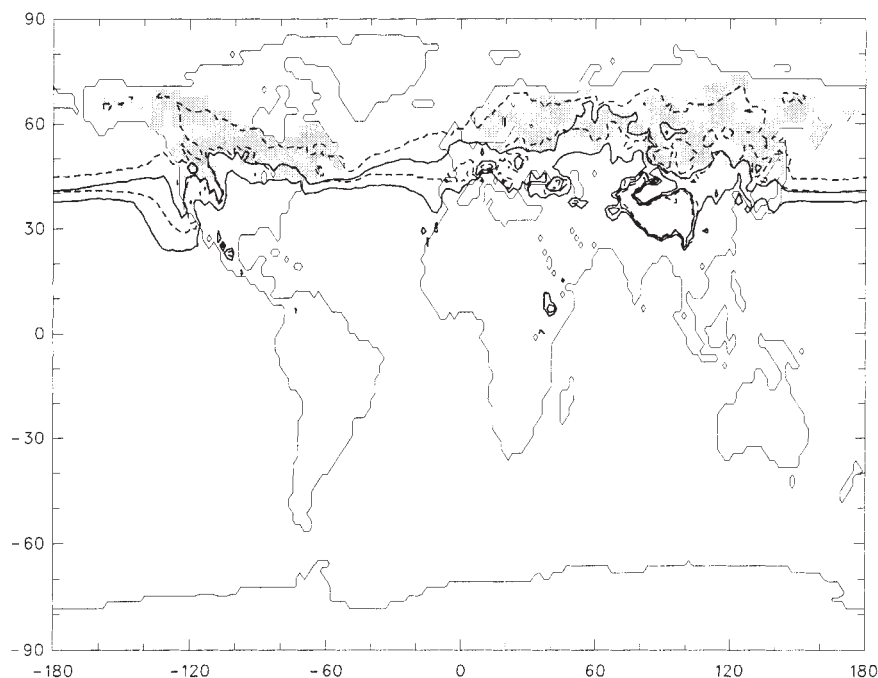


FIG. 3 Relationship between area of forest removed and the 13 °C and 18 °C temperature isotherms for July. Shaded region shows the area of forest vegetation removed for the deforested simulation. (....), Positions of the July 13 °C and July 18 °C isotherms based on observed temperature records. (—), Positions of these isotherms after the forest was removed.

deforested area (Fig. 3). Thus, boreal deforestation may initiate a long-term irreversible feedback in which the forest does not recover and the treeline moves progressively farther south.

The position of the treeline distinguishing boreal forest from tundra vegetation has altered in response to past climate changes^{40–42} and is likely to change with the warmer climate

caused by increased atmospheric CO₂ concentrations^{43–46}. Our studies raise the possibility of considerable climate changes caused merely by redistribution of boreal forest and tundra ecosystems. The decrease in snow-covered land surface albedo caused by northward migration of boreal forest into tundra in response to climate warming may produce further warming. □

Received 4 June; accepted 2 September 1992.

1. Wilson, M. F., Henderson-Sellers, A., Dickinson, R. E. & Kennedy, P. J. *J. Climatol.* **7**, 319–343 (1987).
2. Sato, N. et al., *J. Atmos. Sci.* **46**, 2757–2782 (1989).
3. Sud, Y. C. et al. *Agric. For. Meteorol.* **52**, 133–180 (1990).
4. Dickinson, R. E. & Henderson-Sellers, A. *Q. J. R. Met. Soc.* **114**, 439–462 (1988).
5. Shukla, J., Nobre, C. & Sellers, P. *Science* **247**, 1322–1325 (1990).
6. Nobre, C. A., Sellers, P. J. & Shukla, J. *J. Clim.* **4**, 957–988 (1991).
7. Dickinson, R. E., Meehl, G. A. & Washington, W. M. *Climatic Change* **10**, 241–248 (1987).
8. Ingram, W. J., Wilson, C. A. & Mitchell, J. F. B. *J. geophys. Res.* **94**, 8609–8622 (1989).
9. Meehl, G. A. & Washington, W. M. *Climatic Change* **16**, 283–306 (1990).
10. Federer, C. A. *J. appl. Met.* **7**, 789–795 (1968).
11. Robinson, D. A. & Kukla, G. *J. Clim. appl. Met.* **23**, 1626–1634 (1984); **24**, 402–411 (1985).
12. Otterman, J., Chou, M.-D. & Arking, A. *J. Clim. appl. Met.* **23**, 762–767 (1984).
13. Harvey, L. D. D. *Climatic Change* **13**, 191–224 (1988).
14. Thomas, G. & Rowntree, P. R. *Q. J. R. Met. Soc.* **118**, 469–497 (1992).
15. Pollard, D. & Thompson, S. L. *Description of a Land-Surface-Transfer Model (LSX) as Part of a Global Climate Model, NCAR Tech. Note* (Natl. Center Atmos. Res., Boulder Colorado, 1992).
16. Williamson, D. L., Kiehl, J. T., Ramanathan, V., Dickinson, R. E. & Hack, J. J. *Description of NCAR Community Climate Model (CCM1) NCAR Tech. Note* TN-285 + STR (Natl. Center Atmos. Res., Boulder Colorado, 1987).
17. Slingo, A. & Slingo, J. M. *J. geophys. Res.* **96**, 15341–15357 (1991).
18. Thompson, S. L., Ramaswamy, V. & Covey, C. *J. geophys. Res.* **92**, 10942–10960 (1987).
19. Williamson, D. L. & Rasch, P. J. *Mon. Weath. Rev.* **117**, 102–129 (1989).
20. Rasch, P. J. & Williamson, D. L. *Q. J. R. Met. Soc.* **116**, 1071–1090 (1990).
21. Williamson, D. L. *Tellus* **A42**, 413–428 (1990).
22. Dickinson, R. E., Henderson-Sellers, A., Kennedy, P. J. & Wilson, M. F. *Biosphere–Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model NCAR Tech. TN-275 + STR* (Natl. Center Atmos. Res., Boulder Colorado, 1986).
23. Sellers, P. J., Mintz, Y., Sud, Y. C. & Dalcher, A. *J. Atmos. Sci.* **43**, 505–531 (1986).
24. Covey, C. & Thompson, S. L. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **75**, 331–341 (1989).
25. Dorman, J. L. & Sellers, P. J. *J. appl. Met.* **28**, 833–855 (1989).

26. Olson, J. S., Watts, J. A. & Allison, L. J. *Carbon in Live Vegetation of Major World Ecosystems*, ORNL-5862 (Oak Ridge Natn. Lab., Oak Ridge Tennessee, 1983).
27. Rosencranz, A. & Scott, A. *Nature* **355**, 293–294 (1992).
28. Alexander, R. C. & Mobley, R. L. *Mon. Weath. Rev.* **104**, 143–148 (1976).
29. Lashof, D. A. *Climatic Change* **14**, 213–242 (1989).
30. D'Arrigo, R., Jacoby, G. C. & Fung, I. Y. *Nature* **329**, 321–323 (1987).
31. Bonan, G. B. *Tellus* **B44**, 173–185 (1992).
32. Bonan, G. B. *J. geophys. Res.* **96**, 7301–7312 (1991); 17329–17338 (1991).
33. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. *Climate Change: The IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1990).
34. Post, W. M. *Report of a Workshop on Climate Feedbacks and the Role of Peatlands, Tundra, and Boreal Ecosystems in the Global Carbon Cycle*, ORNL/TM-11457 (Oak Ridge Natn. Lab., Oak Ridge Tennessee, 1990).
35. Bryson, R. A. *Geogr. Bull.* **8**, 228–269 (1966).
36. Larsen, J. A. *The Boreal Ecosystem* (Academic, New York, 1980).
37. Tuhtanen, S. *Acta bot. Fennica* **127**, 1–50 (1984).
38. Oechel, W. C. & Lawrence, W. T. in *Physiological Ecology of North American Plant Communities* (eds Chabot, B. F. & Mooney, H. A.) 66–94 (Chapman & Hall, New York, 1985).
39. Elliot-Fisk, D. L. in *North American Terrestrial Vegetation* (eds Barbour, M. G. & Billings, W. D.) 33–62 (Cambridge Univ. Press, Cambridge, 1988).
40. Ritchie, J. C. *Postglacial Vegetation of Canada* (Cambridge Univ. Press, Cambridge, 1987).
41. Webb, T. III in *Vegetation History* (eds Huntley, B. & Webb, T. III) 385–414 (Kluwer, Boston, 1988).
42. Sirois, L. in *A Systems Analysis of the Global Boreal Forest* (eds Shugart, H. H., Leemans, R. & Bonan, G. B.) 196–215 (Cambridge Univ. Press, Cambridge, 1992).
43. Emanuel, W. R., Shugart, H. H. & Stevenson, M. P. *Climatic Change* **7**, 29–43 (1985).
44. Pastor, J. & Post, W. M. *Nature* **334**, 55–58 (1988).
45. Smith, T. M., Shugart, H. H., Bonan, G. B. & Smith, J. B. *Adv. Ecol. Res.* **22**, 93–116 (1992).
46. Smith, T. M., Leemans, R. & Shugart, H. H. *Climatic Change* **21**, 367–394 (1992).

ACKNOWLEDGEMENTS. The National Center for Atmospheric Research is sponsored by the NSF.

Evidence from Re–Os isotopes for plume–lithosphere mixing in Karoo flood basalt genesis

R. M. Ellam*, R. W. Carlson† & S. B. Shirey†

* Isotope Geology Unit, Scottish Universities' Research and Reactor Centre, East Kilbride, Glasgow G75 0QU, UK

† Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC, 20015, USA

THERE is abundant evidence to link continental flood basalt provinces with mantle plume activity^{1,2}; however, flood basalts often have radiogenic isotope signatures that are outside the range for plume sources as recorded in ocean-island basalts. These signatures are more readily attributed to source regions within the continental lithosphere, and there are strong indications that it is often the mantle lithosphere that contributes isotopically 'enriched' (low-¹⁴³Nd/¹⁴⁴Nd) material^{3,4}. There are objections, however, to models that propose cold, refractory mantle lithosphere as an important source of basaltic magma⁵. The rhenium–osmium system may provide new constraints on the relative importance of plume and sub-continental lithospheric mantle (SCLM) in basalt genesis: samples of SCLM brought to the surface as xenoliths commonly have low ¹⁸⁷Os/¹⁸⁸Os ratios^{6,7}, whereas ocean-island basalts tend to have higher than chondritic ¹⁸⁷Os/¹⁸⁸Os, and old continental crust has even higher (more radiogenic) ratios. Here we report initial ¹⁸⁷Os/¹⁸⁸Os ratios for 190-Myr-old picrite basalts from the Nuanetsi region of the Karoo flood basalt province, which are most readily explained as mixtures between enriched SCLM and sub-lithospheric material, most probably derived from a mantle plume.

Picrite basalts from the Karoo flood basalt province have been the subject of extensive geochemical study^{8–10}, in part because

they are considered to represent primitive magmas similar to parental liquids that evolve to form the abundant, tholeiitic basalts typical of this and other flood basalt provinces^{8,11}.

The radiogenic isotope and incompatible trace element compositions of these rocks have been attributed¹² to mixing between asthenospheric melts that are poor in incompatible trace elements and material derived from the SCLM that is rich in incompatible trace elements. Even small amounts of the lithospheric component are rich enough in most incompatible trace elements to dominate the composition of the hybrid magmas. The incompatible elements, and in particular Nd and Pb isotope ratios, are of limited efficacy as indicators of sub-lithospheric source regions, because they are easily overwhelmed by the SCLM endmember.

By contrast, the Re–Os system does not seem to be strongly influenced by the processes that cause the SCLM to be enriched in incompatible trace elements. From the numerous Re–Os isotope analyses now available for peridotite mantle xenoliths^{6,7} it seems that the SCLM consistently preserves a history of low Re/Os ratio recorded by low ¹⁸⁷Os/¹⁸⁸Os values. Furthermore, even samples that have undergone ancient enrichment in incompatible trace elements retain the low Re/Os signature, so that enriched portions of the SCLM become characterized by both low γ_{Os} and low ϵ_{Nd} .

$$\gamma_{Os} = [((^{187}Os/^{188}Os)_{\text{sample}} / (^{187}Os/^{188}Os)_{\text{mantle}}) - 1] \times 100$$

$$\epsilon_{Nd} = [((^{143}Nd/^{144}Nd)_{\text{sample}} / (^{143}Nd/^{144}Nd)_{\text{mantle}}) - 1] \times 10^4$$

where 'mantle' implies the chondritic estimate for the bulk mantle.

Thus, even portions of the SCLM rich in incompatible trace elements are unlikely to vary greatly in Os concentration or isotopic composition, so ¹⁸⁷Os/¹⁸⁸Os is potentially a much more useful tracer of sub-lithospheric source materials than the isotopic compositions of incompatible elements such as Sr, Nd and Pb. Furthermore, the association of low γ_{Os} with low ϵ_{Nd}

TABLE 1 Re, Os concentrations and Os isotope ratios of Karoo picrite basalts

Sample	N40(I)	N55(I)	N77(I)	N117(I)	N128(I)	N163(I)	N190(I)	N409(I)
Os (pg per g)	148	672	298	1327	341	239	1347	258
$^{187}\text{Os}/^{188}\text{Os}_{(i)}$	0.16024	0.12799	0.13793	0.13290	0.12727	0.14929	0.12776	0.13444
2 s.e.	± 0.00099	± 0.00092	± 0.00105	± 0.00103	± 0.00051	± 0.00113	± 0.00070	± 0.00093
Sample	N55(II)	N77(II)	N117(II)	N128(II)	N163(II)	N409(II)		
Re (pg per g)	127	112	116	55	172	108		
Os (pg per g)	592	381	2130	483	227	263		
$^{187}\text{Re}/^{188}\text{Os}$	0.956	1.424	0.262	0.554	3.523	1.990		
$^{187}\text{Os}/^{188}\text{Os}_{(i)}$	0.13081	0.13451	0.13358	0.13110	0.14936	0.13323		
2S s.e.	± 0.00140	± 0.0043	± 0.00074	± 0.00029	± 0.00065	± 0.00044		
$^{187}\text{Os}/^{188}\text{Os}_{(i)}$	0.12783	0.13007	0.13276	0.12937	0.13837	0.12702		
$\gamma_{\text{Os}} \pm 2 \text{ s.e.}$	$+1.2 \pm 1.1$	$+3.0 \pm 0.3$	$+5.1 \pm 0.6$	$+2.4 \pm 0.2$	$+9.5 \pm 0.4$	$+0.5 \pm 0.3$		
Sample	N40(III)	N128(III)		N163(III)		N190(III)		
Re (pg per g)	164	58		154		121		
Os (pg per g)	135	545		366		1454		
$^{187}\text{Re}/^{188}\text{Os}$	5.885	0.509		2.037		0.400		
$^{187}\text{Os}/^{188}\text{Os}_{(i)}$	0.15822	0.12820		0.14795		0.12504		
2 s.e.	± 0.00036	± 0.00214		± 0.00143		± 0.00019		
$^{187}\text{Os}/^{188}\text{Os}_{(i)}$	0.13986	0.12661		0.14160		0.12379		
$\gamma_{\text{Os}} \pm 2 \text{ s.e.}$	$+10.7 \pm 0.2$	$+0.2 \pm 1.7$		$+12.1 \pm 1.0$		-2.0 ± 0.2		

Data represent the results of separate sample digestions where groups (I), (II) and (III) indicate separate analyses done sequentially. Early analyses of Re (group I) were affected by high and variable blanks associated with the solvent extraction technique then used for Re purification following Os distillation. Consequently, Re concentration data are not reported for these analyses, but the Os data are reported to illustrate the level of reproducibility of these measurements. Re and Os concentrations are expressed as 10^{-12} g per g. Errors on isotope ratios are 2 standard errors of the mean (2 s.e.). For additional isotope and elemental data see refs 9 and 11. Subscript (i) indicates initial isotope ratio.

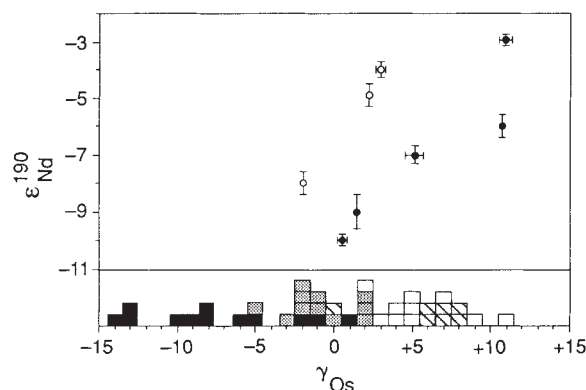
in mantle xenoliths seems to be a unique isotopic signature; if this signature were found in basalts, it would provide compelling evidence that SCLM was involved in their genesis.

To test these assertions, we determined Os isotope compositions and Re and Os concentrations for eight of the Karoo picrite basalts whose elemental and isotopic parameters have previously been well characterized^{10,12}. Three of these samples are from a single area close to the Gomalwe intrusion¹³ and are represented by open symbols in the figures, whereas the remainder (solid symbols) are from more widely spaced localities. The initial Os isotope ratios of the picrite basalts lie within the range recorded for oceanic island basalts (OIB), mid-ocean ridge basalts (MORB) and abyssal peridotites¹⁴⁻¹⁸ (Fig. 1). This contrasts markedly with their Nd isotope ratios ($\epsilon_{\text{Nd}} = -3.3$ to -9.6), which are largely outside the present-day

OIB-MORB range ($\epsilon_{\text{Nd}} = +11$ to -2). The picrite basalts show a broad positive correlation between γ_{Os} and ϵ_{Nd} which improves if the three samples from the Gomalwe area are considered separately, and may even represent two discrete correlations. Complementary relationships are found between γ_{Os} and major and trace element criteria (Fig. 2). Higher γ_{Os} values are associated with higher ϵ_{Nd} , elevated abundances of 'basaltic' major elements (such as Ca, Fe), lower concentrations of incompatible trace elements (such as Ba), and lower ratios of highly incompatible trace elements to moderately incompatible elements (such as Ce/Yb).

The observed correlations are perhaps most simply attributed to mixing between components with differing isotopic composition and elemental abundances. To the low- $^{187}\text{Os}/^{188}\text{Os}$ end of the arrays, the picrite basalts have γ_{Os} values that overlap those

FIG. 1 Plot of ϵ_{Nd} against γ_{Os} , showing correlation for the Karoo picrite basalts. In this and all other figures, samples for which duplicate γ_{Os} values are available are represented by the average values of these duplicates. Error bars (2 standard error) are illustrated when they exceed the symbol size. Errors on concentration determinations are not propagated to the age-corrected initial isotope ratios (i), and therefore represent minimum error estimates. Histogram illustrates published γ_{Os} values for Kaapvaal mantle xenoliths (solid)⁶, mid-ocean ridges basalts and abyssal peridotites (dotted)¹⁸, Gorgona Island komatiites²¹ (cross-hatched) and ocean island basalts (open)¹⁸. Isotope data are corrected to 190 Myr using $\lambda^{147}\text{Sm} = 6.54 \times 10^{-12} \text{ yr}^{-1}$ where λ is the decay constant and the most recent experimentally determined²⁵ value of $\lambda^{187}\text{Re} = 1.639 \times 10^{-11} \text{ yr}^{-1}$. METHODS. Os and Re were separated from acid digestions of samples (2-4 g) of whole-rock powder using distillation²⁶ and anion exchange²⁷ respectively. Total procedure blanks ranged from 5 to 10 pg for Os and averaged 20 pg for Re. Measurements were made by negative thermal ionization mass spectrometry²⁸ on a 15-inch-radius mass spectrometer at the Department of Terrestrial Magnetism, Carnegie Institution. Repeat analyses of standard solutions show reproducibility of better than 2% and 0.5% for Re and Os concentrations, respectively. But as seen in the results, powder heterogeneity of the samples can contribute to the variation of replicate measurements beyond these limits. Here we have broken with the convention of presenting $^{187}\text{Os}/^{186}\text{Os}$ in favour of $^{187}\text{Os}/^{188}\text{Os}$. ^{188}Os can be measured with greater precision than ^{186}Os and has fewer isobaric interferences. Furthermore, there is a possible contribution to ^{186}Os from



the decay of ^{190}Pt . Although this is not considered to be very significant here, the improved measurement precision afforded by negative thermal ionization mass spectrometry over previous methods will exacerbate problems with ^{186}Os normalization, making a switch to $^{187}\text{Os}/^{188}\text{Os}$ desirable. γ_{Os} values are not affected by this change, and the 'mantle' $^{187}\text{Os}/^{188}\text{Os}$ value adopted here (0.12776) is consistent with $^{187}\text{Os}/^{186}\text{Os} = 1.060$ at the present day.

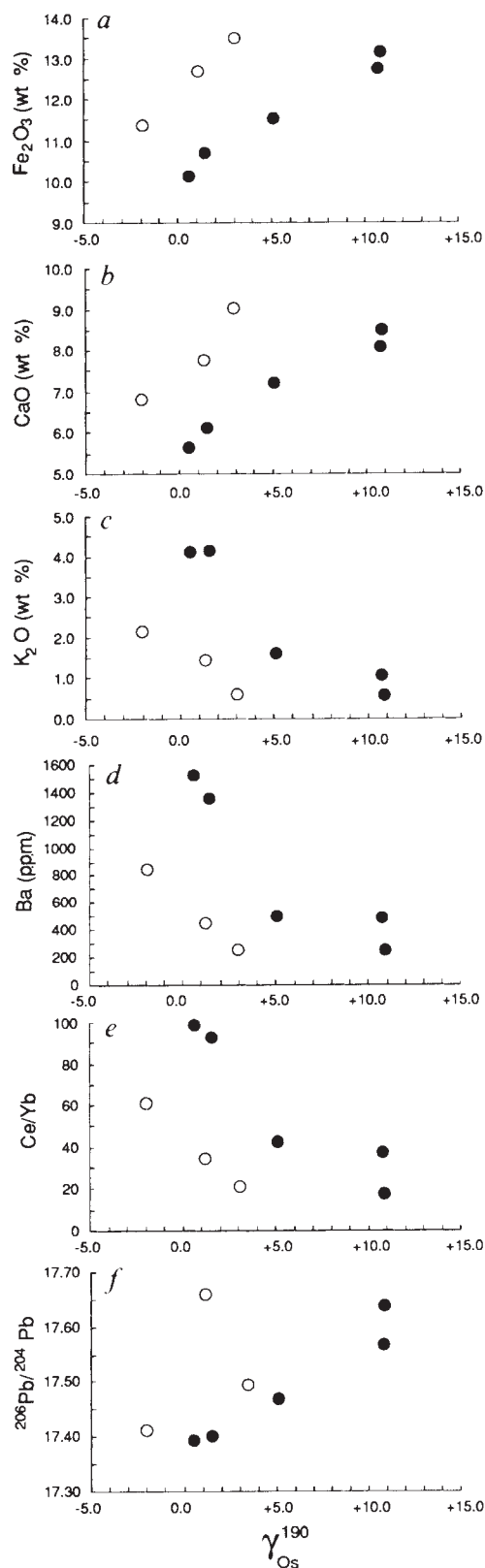


FIG. 2 Plots of γ_{Os} against geochemical parameters. Broad correlations exist between γ_{Os} and major (a–c) and trace (d, e) element parameters, and with the exception of one sample (for which we offer no explanation), a strong relationship between γ_{Os} and $^{206}\text{Pb}/^{204}\text{Pb}$ (f) is also suggested.

of MORB and approach the characteristic negative γ_{Os} of lithospheric mantle xenoliths. These low γ_{Os} values are associated with low ϵ_{Nd} ratios (Fig. 1), ruling out MORB-like material as a possible endmember but consistent with a low-Sm/Nd lithospheric endmember.

Continental-crust rocks often carry a low ϵ_{Nd} signature, but are unlikely to have low γ_{Os} , because Re/Os ratios are expected to be high in evolved crustal lithologies¹⁹. Even mafic crust, for which ϵ_{Nd} could range from high to low values, would still have a fairly high γ_{Os} . Crustal contamination is therefore unlikely to generate both low ϵ_{Nd} and low γ_{Os} .

Rhenium concentrations for the low- γ_{Os} picrite basalts are significantly below those typical of asthenospheric melts^{18,20,21} including MORB²², and this suggests that a Re-depleted component, such as SCLM, is involved. Moreover, the low- γ_{Os} , supposed lithospheric, endmember is also associated with enhanced concentrations of incompatible trace element and lowered abundances of 'basaltic' major elements. The combination of 'basaltic' major element depletion and incompatible trace element enrichment is a common feature of mantle xenoliths and therefore strongly suggestive of SCLM source material. Furthermore, with the exception of one sample (N55, for which we can offer no explanation), there is a strong correlation between γ_{Os} and $^{206}\text{Pb}/^{204}\text{Pb}$, suggesting a low- $^{206}\text{Pb}/^{204}\text{Pb}$ component similar to the source of group II kimberlites²³ and some lamproites²⁴ for which a SCLM origin is favoured. Thus, on several grounds, the most likely endmember to carry the combined low- γ_{Os} , low- ϵ_{Nd} signature is Re-poor, lithophile-element-enriched SCLM.

The highest initial $^{187}\text{Os}/^{188}\text{Os}$ recorded in the picrite basalts are within the range recorded for OIB, and therefore consistent with a sub-lithospheric, plume-related, origin. The precise nature of the presumed sub-lithospheric endmember remains more enigmatic. It might be imprudent to attach strong significance to compositional variations that are based on relatively few data, particularly when the trends identified occur in diagrams for which mixing curves are not necessarily linear (Figs 1–3). Unfortunately, in this case, diagrams of $^{187}\text{Os}/^{188}\text{Os}$ against $1/\text{Os}$ and so on do not help to constrain mixing hypotheses further, presumably because concentrations of Os are affected by its compatible behaviour during magma generation and differentiation processes. Tentative conclusions about the likely plume endmember may, however, be drawn.

Mixing between a high-degree partial melt of a plume and a low-degree, incompatible-element-rich, osmium-poor melt of the SCLM will result in strongly convex-downward mixing curves in Fig. 1 (see Fig. 3). Because the picrite basalts show only a shallow correlation between Nd and Os isotope variation, the data neither require nor rule out the hyperbolic mixing curves (Fig. 3). The fact that incompatible element concentrations increase with decreasing ϵ_{Nd} (ref. 12) and γ_{Os} (for example, Fig. 2d) is at least consistent with the illustrated mixing trends. It is worth emphasizing that the proposed plume endmember is relatively poor in incompatible trace elements compared with the lithospheric component. Previous work¹² has used this characteristic to suggest a sub-lithospheric component that is more akin to MORB than to OIB, whereas the Os isotope data seem to favour an OIB-like component. It is important, however, not to confuse the trace element and isotope evidence. The isotopic distinction between MORB and OIB is a source distinction, whereas the trace element differences may in part reflect different conditions of melting. Ocean island and continental flood basalt occurrences represent different expressions of plume activity. In particular, very high degrees of melting are most probably required to generate flood basalts with MgO contents of ~15 wt%. It is therefore expected that the Karoo plume component would carry an OIB-like isotope signature, but would have lower abundances of incompatible trace element than typical OIB.

Finally, we should contrast the broad correlations between

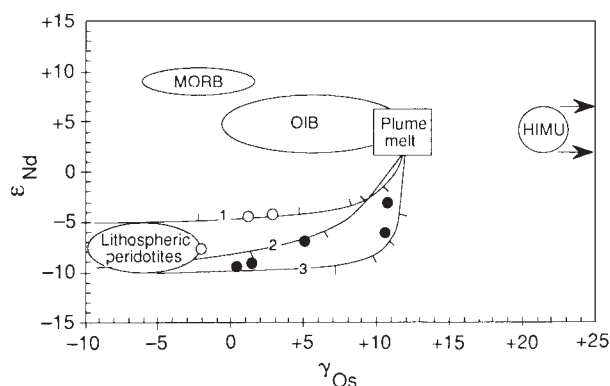


FIG. 3 Proposed mixing relationships illustrated on a plot of ϵ_{Nd} against γ_{Os} . Isotope data are corrected to the appropriate age for each sample (or field). Symbols for picrite basalts as in Fig. 2. The most important point is that rather flat-lying arrays can be generated by feasible mixing models despite the involvement of a plume endmember with fairly high ϵ_{Nd} . Fields for MORB (largely based on abyssal peridotite data), OIB and lithospheric peridotites are illustrated to emphasize the likely sub-lithospheric and lithospheric endmembers, respectively; thus, although lithospheric peridotites extend to positive ϵ_{Nd} values, we illustrate a range that is consistent with the required low- ϵ_{Nd} endmember. Also shown is a field for the HIMU²⁹ subset of OIB showing highly elevated γ_{Os} ³⁰. Mixing curves are ticked to represent 20% increments, and were calculated using the following parameters. Plume endmember: γ_{Os} , +12; Os, 1,000 pg per g; ϵ_{Nd} , +3; Nd, 10 μg per g. Mantle lithosphere endmember: model 1: γ_{Os} , -20; Os, 200 pg per g; ϵ_{Nd} , -5; Nd, 50 μg per g; model 2: γ_{Os} , -20; Os, 200 pg per g, ϵ_{Nd} , -10; Nd, 50 μg per g; model 3: γ_{Os} , -20; Os, 500 pg per g; ϵ_{Nd} , -10; Nd, 50 μg per g. Os and Nd concentrations were chosen to represent asthenospheric melts and small-volume lithospheric melts, such as lamproite or kimberlite⁶, respectively.

γ_{Os} and elemental abundances with the highly significant correlations observed previously between the ϵ_{Nd} values and the concentrations of several major and trace elements for these samples^{10,12}. Several explanations could be proposed, including variation in Os/Nd ratio of the endmembers, existence of a third endmember distinguished only by Os isotope composition, and open-system behaviour of Os, but all are speculative. Although our data set shows that such complexity is necessary, it cannot differentiate between a variety of possibilities, nor constrain further the likely source components.

The new Os isotope data strongly indicate that SCLM has a role in the petrogenesis of Karoo flood basalts. Above all, the association of low γ_{Os} with low ϵ_{Nd} seems to confirm that a trace-element-rich portion of the mantle lithosphere is involved in the formation of these basalts. Endmember γ_{Os} compositions are not well constrained and mixing models are therefore schematic. A lithospheric component based on the probable low Os content of low-melt-fraction magmas such as kimberlite or lamproite is, however, broadly consistent with the new data, and can support the large proportions (up to 50%) of lithospheric component inferred from Nd isotope and elemental data for the most incompatible-element-enriched picrite basalts¹².

The Os isotope data apparently also support the idea that OIB-like plume material has a role in Karoo petrogenesis. The role of plume-lithosphere mixing in the petrogenesis of flood basalts is thus firmly substantiated by the new Os isotope data. This is an important observation: first because it begins to reconcile the geochemical constraints on magma source region with geophysical models of large-scale plume-related magma generation, and second because it confirms the importance of the SCLM as a substantial contributory source region for basaltic magmatism. □

4. Hawkesworth, C. J., Marsh, J. S., Duncan, A. R., Erlank, A. J. & Norry, M. J. *Geol. Soc. S. Afr. spec. Publ.* **13**, 342-354 (1984).
5. Arndt, N. T. & Christensen, U. *J. geophys. Res.* **97**, 10967-10981.
6. Walker, R. J., Carlson, R. W., Shirey, S. B. & Boyd, F. R. *Geochim. cosmochim. Acta* **53**, 1583-1595 (1989).
7. Pearson, D. G. et al. *Ext. Abstr. 5th int. Kimberlite Conf., Araxá, Brazil, C.P.R.M. spec. Publ.* 2/91, 329-331 (Brazilia, 1991).
8. Cox, K. G. & Jamieson, B. G. *J. Petrol.* **15**, 269-301 (1974).
9. Cox, K. G., Duncan, A. R., Bristow, J. W., Taylor, S. R. & Erlank, A. J. *Geol. Soc. S. Afr. spec. Publ.* **13**, 149-169 (1984).
10. Ellam, R. M. & Cox, K. G. *Earth planet. Sci. Lett.* **92**, 207-218 (1989).
11. Bristow, J. W., Allsopp, H. L., Erlank, A. J., Marsh, J. S. & Armstrong, R. A. *Geol. Soc. S. Afr. spec. Publ.* **13**, 295-329 (1984).
12. Ellam, R. M. & Cox, K. G. *Earth planet. Sci. Lett.* **105**, 330-342 (1991).
13. Jamieson, B. G. thesis, Univ. of Edinburgh (1969).
14. Luck, J.-M., Pegram, W. J. & Allègre, C. J. *Chem. Geol.* **70**, 54 (1988).
15. Pegram, W. J., Luck, J.-M. & Allègre, C. J. *Chem. Geol.* **70**, 55 (1988).
16. Pegram, W. J., Luck, J.-M. & Allègre, C. J. *EOS* **69**, 510 (1988).
17. Pegram, W. J. & Allègre, C. J. *Terra Abstr.* **1**, 343 (1989).
18. Martin, C. E. *Geochim. cosmochim. Acta* **55**, 1421-1434 (1991).
19. Allègre, C. J. & Luck, J.-M. *Earth planet. Sci. Lett.* **48**, 148-154 (1980).
20. Walker, R. J., Shirey, S. B. & Stecher, O. *Earth planet. Sci. Lett.* **87**, 1-12 (1988).
21. Walker, R. J., Echeverria, L. M., Shirey, S. B. & Horan, M. F. *Contrib. Miner. Petrol.* **107**, 150-162 (1991).
22. Hertogen, J., Janssens, M.-J. & Palme, H. *Geochim. cosmochim. Acta* **44**, 2125-2143 (1980).
23. Smith, C. B. *Nature* **304**, 51-54 (1983).
24. Fraser, K. J., Hawkesworth, C. J., Erlank, A. J., Mitchell, R. H. & Scott-Smith, B. H. *Earth planet. Sci. Lett.* **76**, 57-70 (1985).
25. Lindner, M., Leich, D. A., Russ, G. P., Bazan, J. M. & Borg, R. J. *Geochim. cosmochim. Acta* **53**, 1597-1606 (1989).
26. Walker, R. J. *anal. Chem.* **60**, 1231-1234 (1988).
27. Morgan, J. W., Golightly, G. W. & Dorrzapf, A. F. Jr. *Talanta* **38**, 259-265 (1991).
28. Creaser, R., Papanastassiou, D. A. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **55**, 397-401 (1991).
29. Zindler, A. & Hart, S. R. *Rev. Earth planet. Sci.* **14**, 493-571 (1986).
30. Hauri, E. H. & Hart, S. R. *EOS* **72**, 507 (1991).

ACKNOWLEDGEMENTS We thank K. G. Cox for making the Nuanetsi samples available to us and for discussions. This work was supported by NSF (R.W.C., S.B.S.) and NERC (R.M.E.).

Saharan exploitation of plants 8,000 years BP

Fred Wendorf*, Angela E. Close*, Romuald Schild†, Krystyna Wasylikowa‡, Rupert A. Housley§, Jack R. Harlan|| & Halina Królik†

*Department of Anthropology, Southern Methodist University, Dallas, Texas 75275-0336, USA

†Instytut Archeologii i Etnologii, Polskiej Akademii Nauk (d. Instytut Historii Kultury Materialnej), Al. Solidarności 105, 00-140 Warszawa, Poland

‡Instytut Botaniki, Polskiej Akademii Nauk, Lubicz 46, 31-512 Kraków, Poland

§Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, 6 Keble Road, Oxford OX1 3QJ, UK

||1016 North Haggar Street, New Orleans, Louisiana 70119, USA

SORGHUM and millets are among the world's most important food crops and, for the inhabitants of the semi-arid tropics, they are the main sources of protein and energy. Little is known about the history of these crops; their domestication is thought to have occurred in the African savannah, but the date and precise location are unknown^{1,2}. Excavations at an early Holocene archaeological site in southernmost Egypt, 100 km west of Abu Simbel, have yielded hundreds of carbonized seeds of sorghum and millets, with consistent radiocarbon dates of 8,000 years before present (BP), thus providing the earliest evidence for the use of these plants. They are morphologically wild, but the lipid fraction of the sorghum grains shows a closer relationship to domesticated than to wild varieties. Whatever their domestic status, the use of these plants 8,000 years ago suggests that the African plant-food complex developed independently of the Levantine wheat and barley complex.

Recent excavations at an early Holocene archaeological site in the Egyptian Sahara have yielded several thousand carbonized remains of plant foods. They represent more than 40 different plants, including sorghum, millets, legumes, fruits, nuts and tubers that are presumed to be wild (although there are some

Received 14 May; accepted 26 August 1992.

1. Morgan, W. J. in *The Sea*, Vol. 7 (ed. Emiliani, C.) (Wiley, New York, 1981).
2. White, R. S. & McKenzie, D. *J. geophys. Res.* **94**, 7685-7730 (1989).
3. Carlson, R. W. *Geochim. cosmochim. Acta* **48**, 2357-2372 (1984).

suggestions of cultivation in the sorghum). Accelerator mass-spectrometric ^{14}C dates on 11 of the seeds all fall within one standard deviation of 8,000 yr BP. The early Holocene plants were found at Site E-75-6, at Nabta Playa, near the Egyptian-Sudanese border.

The site had been excavated in 1975 and 1977, when almost 1,000 m² were cleared, revealing cultural deposits belonging to two phases of Early Neolithic, of which the earlier was dated to 8,800–8,500 yr BP. The upper horizon of the site showed an arrangement of two rows of features, including the shallow basin-floors of huts, hearths, storage pits and walk-in wells; it gave several ^{14}C dates on charcoal of 8,100–8,000 yr BP.

The site was re-excavated in 1990 and 1991, in an area south-east of the earlier excavations (in the direction of the two lines of features) and sealed beneath playa clays, unlike the areas of the 1970s. Removal of the clays uncovered four houses, two oval and two round, and twelve pits within a 10×15 m area; only two quadrants of each house have been excavated. The two oval houses were large (8.3×4.5 m and 7×2.5 m), shallow basins. The floor of one of them had traces of six hearths (<1 m² in area) and, separate from them, at least 74 small hemispherical depressions, or cooking-holes, usually 10–20 cm in diameter (Fig. 1). The excavated portion of the other oval house had four hearths and 18 cooking-holes. The shallow, circular houses are about 4 m in diameter. One had three hearths and seven cooking-holes plus an irregular bell-shaped storage pit, and the other had a single hearth and 23 cooking-holes. Two pits crossed the edge of one of the oval houses, indicating more than one occupation of the site; however, the accelerator dates indicate that all the houses and pits were used around 8,000 yr BP (Table 1).

In all the houses, a burned, ashy sediment had been piled up around some of the cooking-holes while they held containers, and this sediment was particularly rich in plant remains. This

TABLE 1 Accelerator mass-spectrometric ^{14}C dates on charred plant remains from Site E-75-6, Nabta Playa, southern Egypt

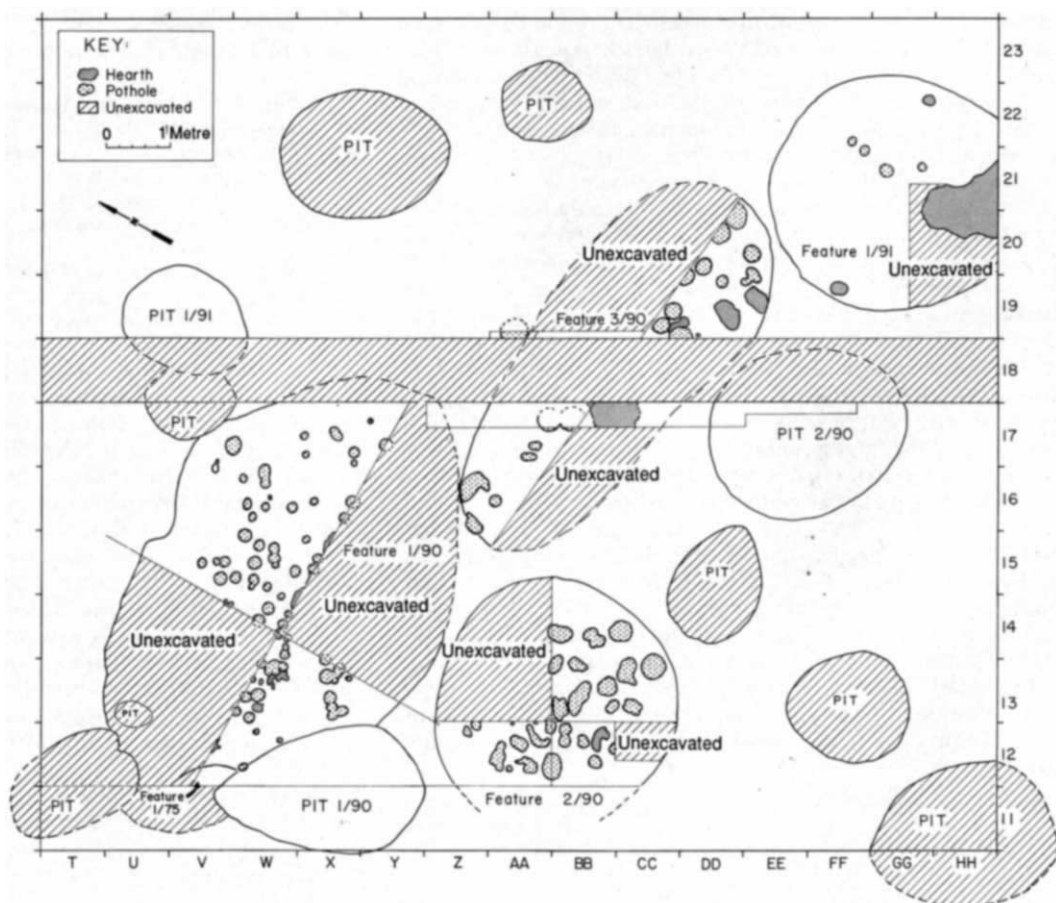
Laboratory number	Plant type	Location	^{13}C (‰)	Uncalibrated ^{14}C date BP
OxA-3214	Cruciferae	Feature 1/90	NA	8,080 ± 110
OxA-3215	Leguminosae	Feature 1/90	NA	8,095 ± 120
OxA-3216	Sorghum	Feature 1/91	-9.6	7,960 ± 100
OxA-3217	Sorghum	Feature 1/90	-9.1	8,020 ± 160
OxA-3218	Zizyphus	Feature 1/90	-24.3	8,050 ± 130
OxA-3219	Sorghum	Feature 3/90	-9.5	7,950 ± 160
OxA-3220	Zizyphus	Pit 1/90	-23.1	8,025 ± 120
OxA-3221	Sorghum	Feature 1/90	-9.1	7,980 ± 110
OxA-3222	Sorghum	Feature 2/90	-9.4	8,060 ± 120
OxA-3484	Panicum-type	Feature 2/90	-8.9	7,950 ± 90
OxA-3485	Zizyphus	Pit (BB/10)	-17.2	7,980 ± 95

Dates are expressed in radiocarbon years before the present (AD 1950), using the half-life of 5,568 years. Isotope fractionation has been corrected for by measuring the ^{13}C (to within ±0.5–1.0‰ relative to the standard rock, PDB), except for OxA-3214 and OxA-3215, for which an assumed value of -26‰ was used. All age determinations have been normalized to -25‰ in accordance with international convention¹². Errors are quoted ±1 s.d. and represent the total error of the accelerator mass-spectrometric system, including the sample chemistry. Chemical pretreatment was as described in ref. 13; ^{14}C measurement as ref. 14. Locations as Fig. 1. NA, not applicable.

suggests that containers of food were placed in the cooking-holes and that hot ash was piled around them to cook the contents, which sometimes boiled over or fell into the ash, leading to carbonization and preservation.

The plant taxa identified thus far are listed in Table 2. The collections are dominated by seeds probably of the mustard and caper families, *Zizyphus* stones, grass grains, and seeds of leguminous plants; the unidentified material consists of many examples of a few types. There is very marked patterning in the association of different taxa with different cooking-holes. Some

FIG. 1 Detailed plan of the areas of Site E-75-6, Nabta Playa, excavated in 1990 and 1991, showing the outlines of the house floors (features) with hearths and cooking-holes (or potholes).



holes yielded predominantly Leguminosae (sometimes a single taxon, sometimes several taxa); some yielded Gramineae (including sorghums and millets); some Cruciferae, Cucurbitaceae, Cyperaceae or Boraginaceae (*Arnebia*); some *Zizyphus* fruits; and some yielded mixtures in differing proportions (Fig. 2).

Several grasses were collected, as they still are today by Saharan nomads^{3,4}. These included primarily *Sorghum* (which biometrically represents a single, fairly homogeneous population), and the smaller-grained grasses resembling *Panicum*, *Digitaria*, *Brachiaria*, *Urochloa*, *Echinochloa* and *Setaria*. Several species of Leguminosae were also collected for their seeds, and probably for tubers as well. The abundance and diversity of the plants indicate that the local vegetation was rich in species and, in some seasons, relatively luxuriant.

All the plants represented are members of the natural sub-desertic or sahelian flora. The sorghum and millets correspond in size and structure to modern wild varieties of these plants and thus seem to be parts of an economy based on gathering wild plant foods. But preliminary investigation by infrared spectroscopy of the lipids in the sorghum grains suggest the possibility of some cultivation⁵. Five archaeological specimens from four features were compared with specimens (from the herbarium at Kew) of three cultivated races of *Sorghum bicolor* and four wild species of *Sorghum*: *S. arundinaceum*, *S. aethiopicum*, *S. verticilli florum* and *S. virgatum*. The spectrographic results for the archaeological specimens are all very similar, and are closer to the cultivated specimens than to the wild, particularly in the hexane extracts. Considerably more work will be required to investigate this possibility, but it is particularly intriguing as it has been suggested that sorghum and millet were domesticated somewhere in the African savannah zone^{1,2}.

The wild relatives of most of the plants of the African savannah complex grow today in the savannah and it has been assumed that they were domesticated somewhere within that zone². There are several theories of where and when this occurred⁶⁻⁸, but there has been no direct archaeological evidence. This absence results primarily from the assumption that the domestication of African plants was a late and secondary development, which began well after wheat and barley agriculture was established in Egypt, about 6,400 yr BP⁹. The problem

TABLE 2 Plant taxa identified from Site E-75-6

Trees
<i>Zizyphus</i> spp.
Grasses (subfamily Panicoideae, tribe Paniceae)
<i>Echinochloa</i> -type
<i>Digitaria</i> -type
<i>Urochloa/Brachiaria</i> -type
<i>Panicum</i> -type
<i>Setaria</i> -type
(subfamily Panicoideae, tribe Andropogoneae)
<i>Sorghum</i> spp.
Other herbs or shrublets
<i>Cyperus/Fuirena</i> -type
<i>Scirpus/Schoenoplectus</i> -type
<i>Rumex</i> -type
Cucurbitaceae
<i>Arnebia</i> -type
Cruciferae (tribe Brassiceae)
Unidentified
Leguminosae, several types
Chenopodiaceae?
Malvaceae?
Capparidaceae?
Tubers (unidentified parenchymatous tissues)
Several unidentified fruits and seeds

is compounded by the poor preservation of organic (especially plant) materials in the recent savannah.

The history of plant domestication in south-western Asia is comparatively well known. Cereals and pulses began to be used intensively in many areas around 12,000 yr BP¹⁰, and shortly after 10,000 yr BP domestic varieties of peas, lentils, barley and emmer wheat occur in several sites¹¹. The subsequent transformation to a food-producing economy was rapid but not uniform; some localities continued to use wild plants long after the domestic forms were used in adjacent areas, but by 8,000 yr BP domestic wheat and barley were in general use throughout the Levant. There is no evidence, however, that these Levantine plants were introduced into Africa until much later: in Egypt, the earliest known wheat and barley are in the Fayum Neolithic, with radiocarbon dates of about 6,400 yr BP⁹.

FIG. 2 Charred plant remains from Site E-75-6. a-c, Grains of *Sorghum* sp. showing the ventral (a) and dorsal (b, c) surfaces. d, Grain of *Panicum*-type. e, Grain of *Echinochloa*-type. f, Complete fruit-stone of *Zizyphus* sp. g, Inner surface of a fragment of a *Zizyphus* fruit-stone, showing two cells. h, Seed of *Zizyphus* sp. Each scale bar, 1 mm. (Photographs by A. Pachonski.)



Today, the northern boundary of the savannah runs east-west, just north of the latitude of Khartoum in central Sudan, and generally coincides with the 100 mm isohyet¹. The northward movement and intensification of the monsoons during the early Holocene were apparently sufficient to shift this boundary 600–700 km northwards into southern Egypt. The evidence for the intensive use of sorghum and millets some 8,000 years ago indicates that such use first occurred much earlier than previously thought, and took place in what is today a rainless desert. It is a short step from intensive use of wild plants, to planting and protecting those plants, and thus to the genetic changes involved in domestication. □

Received 15 May; accepted 4 September 1992.

1. Harlan, J. R. *Crops and Man* (Am. Soc. Agronomy, Crop. Sci. Soc. of Am., Madison, 1975).
2. Harlan, J. R., De Wet, J. M. J. & Stemler, A. B. L. in *Origins of African Plant Domestication* (eds Harlan, J. R., De Wet, J. M. J. & Stemler, A. B. L.) 3–19 (Mouton, The Hague, 1976).
3. Harlan, J. R. *Sahara* 2, 69–74 (1989).
4. Harlan, J. R. in *Foraging and Farming: the Evolution of Plant Exploitation* (eds Harris, D. R. & Hillman, G. C.) 79–98 (Unwin Hyman, London, 1989).
5. Wasylkowska, K. et al. in *Food, Metal and Towns in African Archaeology* (eds Ucko, P. & Shaw, T.) (Routledge, London, in the press).
6. Chevalier, A. *Mém. Soc. Biogéog.* 6, 307–322 (1938).
7. Vignier, P. *Revue int. Bot. appl. Agric. trop.* 25, 163–230 (1945).
8. Murdock, G. P. *Africa: its Peoples and Their Cultural History* (McGraw-Hill, New York, 1959).
9. Hassan, F. A. *J. Wild Prehist.* 2, 135–185 (1988).
10. Bar-Yosef, O. & Belfer-Cohen, A. *J. Wild Prehist.* 3, 447–498 (1989).
11. Moore, A. M. T. *Adv. Wild Archaeol.* 4, 1–69 (1985).
12. Stuiver, M. & Polach, H. A. *Radiocarbon* 19, 355–363 (1977).
13. Hedges, R. E. M., Law, I. A., Bronk, C. R. & Housley, R. A. *Archaeometry* 31, 99–113 (1989).
14. Bronk, C. R. & Hedges, R. E. M. *Nucl. Instrum. Meth.* B29, 45 (1987).

Eusociality in Australian gall thrips

Bernard J. Crespi

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

STUDIES of the role of haplodiploidy in the evolution of eusociality have been limited to the Hymenoptera, the only insects known to exhibit both reproductive castes and the haplodiploid genetic system¹. Because aculeate Hymenoptera share many other traits that may affect sociality, such as provisioning at nests, powerful flight, mandibulate mouthparts, and stings, it has been difficult to separate the effects of haplodiploidy from other characteristics of this taxonomic group^{2,3}. Here I report the presence of eusociality in a second haplodiploid insect taxon, the order Thysanoptera. Subfertile 'soldier' adults of the Australian gall thrips *Oncothrips tepperi* Karny and *O. habrus* Mound defend the gall containing their mother and siblings from invasion and takeover by inquiline thrips species and other insect invaders. Australian gall thrips provide remarkable new opportunities for analysing the causes of the evolution of eusociality.

Oncothrips habrus and *O. tepperi* form galls on phyllodes (petioles modified as leaves) of *Acacia melvillei* and *A. oswaldii*, respectively, in the semiarid and arid zones of Australia^{4–6}. Galls are produced in spring by single, inseminated, macropterous (fully winged) females inserting their mouthparts into young, growing phyllode tissue^{4,5}. During the period from gall initiation until closure, foundresses fight each other with their armed forelegs over gall ownership⁴. After a foundress is enclosed by the gall, she feeds on the plant tissue and oviposits on the inner gall surface, and her offspring feed, develop and eclose inside.

The adult offspring of foundresses comprise two discrete morphological types: macropterous individuals and micropterous (short-winged) individuals of both sexes, which have larger forelegs, shorter antennae and paler cuticles than fully winged adults⁴ (Fig. 1). Micropterous adults eclose earlier than macropterous ones, so that for a period of at least two weeks, and probably longer, galls contain the macropterous foundress, from 3 to 20 micropterous females, from 0 to 5

TABLE 1 Experimental test for division of labour in *O. habrus* and *O. tepperi*

<i>Oncothrips habrus</i> Experiment 1						<i>Oncothrips habrus</i> Experiment 2					
Attacked			Stayed in gall			Attacked			Stayed in gall		
WR	F	M	WR	F	M	WR	F	M	WR	F	M
4	0	0	7	1	0	2	0	0	8	1	0
4	0	0	4	1	0	2	0	0	15	1	1
4	0	0	2	1	0	2	0	0	9	1	2
4	0	0	9	1	0	2	0	0	12	1	0
4	0	0	4	1	0	2	0	0	7	1	0
4	0	0	3	1	0	2	0	0	8	1	2
4	0	0	5	1	1	2	0	0	7	1	0
4	0	0	9	1	0	2	0	0	7	1	1
4	0	0	10	1	2	1	1	0	3	0	2
4	0	0	6	1	1						
4	0	0	4	1	1						
4	0	0	6	1	1						
4	0	0	6	1	1						
4	0	0	11	1	2						
4	0	0	1	1	0						
4	0	0	9	1	0						
4	0	0	5	1	0						
4	0	0	7	X	1						
3	0	1	9	1	2						
3	0	0	3	1	0						
2	0	0	6	1	0						
2	0	0	5	1	0						
2	0	0	7	1	0						
1	0	0	4	1	2						
1	0	1	2	1	8						

Z (WR versus F females)=3.59
P=0.0006

Z (WR versus F females)=0.23
P=0.388

Oncothrips habrus Experiment 3

Oncothrips tepperi

Emerged			Stayed in gall			Emerged			Stayed in gall		
WR	F	M	WR	F	M	WR	F	M	WR	F	M
4	0	0	3	1	0	2	0	0	5	1	0
4	0	0	4	1	1	2	0	0	1	1	1
4	0	0	6	1	1	2	0	0	3	1	1
4	0	0	3	1	0	2	0	0	5	1	0
3	0	0	2	1	0	2	0	1	2	1	1
3	0	0	3	1	0	2	0	0	2	1	0
3	0	0	4	1	0	2	0	0	1	1	1
3	0	0	2	1	0	2	0	0	6	1	1
2	0	0	1	1	0	2	0	0	4	1	1
2	0	0	4	1	0	2	0	0	5	1	0
2	0	0	4	1	0	2	0	0	5	1	1
1	0	0	3	X	0	1	1	0	3	0	1
0	1	0	0	0	1	1	1	0	3	0	0
0	1	0	0	0	0	1	1	0	4	0	0
						1	0	0	6	1	0
						1	0	0	2	1	1
						1	0	0	1	1	1

Z (WR versus F females)=2.62
P=0.0129

Z (WR versus F females)=1.195
P=0.196

Differences between foundresses and micropterous adults in propensity to attack invaders were investigated by cutting holes in the sides of galls containing a foundress, micropterous adults, larvae and eggs, and either (1) removing the first two or four adults that came partially out of the hole, or (2) holding a *Koptothrips* in front of the adults that appear at the hole and come partially out of it, and removing the first two or four adults to attack the *Koptothrips*. Because adults that emerged from the gall nearly always attacked *Koptothrips* held in front of them, emergences and attacks should provide similar estimates of tendency to defend the gall. Each trial lasted five minutes or until the desired number of adults had emerged or attacked. WR, micropterous females; F, foundresses; X, foundresses that were dead before the experiment; M, micropterous males. Attack or emergence probabilities of micropterous females and foundresses were compared with the Mantel-Haenszel test for combining 2 × 2 contingency tables, with adjustment for small sample sizes. All trials with the foundress and one or more live micropterous adults were included in the analyses. By Fisher's combining test, $P < 0.01$ overall for experiments 1–3.

micropterous males, several pupae destined to become micropterous adults, dozens of first- and second-instar larvae destined to become macropterous adults, and eggs (Fig. 1). Macropterous adults eclose between March and June, and in *O. tepperi* the sex ratio of macropterous adults after eclosion is 0.148 male (s.d. = 0.098, $n = 15$ galls), probably as a result of local mate competition⁴. Macropterous adults disperse from the galls, apparently between June and September but, at least in *O. tepperi*, breeding may continue in the galls for a second generation after dispersal of all but one macropterous female⁴.

The presence of enlarged, armed forelegs on micropterous adults of *O. tepperi* and *O. habrus* suggests that these adults are specialized for fighting⁴. Against what do micropterous *Oncothrips* fight, and in what context? When small holes are cut in galls of *O. tepperi* or *O. habrus*, several micropterous adults usually come to the opening and remain just inside it, sometimes staying immobile with their antennae and heads facing out of the hole, and sometimes moving around slowly and protruding their abdomens out of the opening. Adults at such gall openings attacked adults of *Koptothrips* spp. ('inquiline' thrips that invade galls of other species, kill the gall-formers, and breed inside^{4,5}) ($n = 33$ in *O. tepperi*, 18 in *O. habrus*), lepidopteran larvae ($n = 25$ in *O. tepperi*) and *Iridomyrmex humilis* ants ($n = 4$ in *O. tepperi*), usually immediately after these were held in front of them. In *O. tepperi*, foundresses attacked *Koptothrips* twice and lepidopteran larvae three times, and all other attacks were by micropterous adults. In *O. habrus*, all attacks except one by a foundress on a *Koptothrips* were by micropterous adults, and one attack on a *Koptothrips* led to the micropterous female being grasped and killed. Adults at gall openings did not attack micropterous conspecifics from other galls ($n = 20$ in *O. tepperi*, 9 in *O. habrus*), or macropterous conspecifics from other galls held in front of them ($n = 8$ in *O. tepperi*, 9 in *O. habrus*, with one short attack in *O. tepperi*, followed immediately by five replications without attacks). These experiments indicate that adults of *O. tepperi* and *O. habrus* attack insects of other species and not their own.

The common finding of *Koptothrips* breeding in field-collected galls of *O. tepperi* and *O. habrus* containing dead *Oncothrips* suggested that these usurpers represent a major threat. To clarify the nature of interactions between *Oncothrips* and *Koptothrips*, observations of interactions between adults were conducted in closed half-galls. The observations revealed that adults of *O. tepperi* attempt to use their enlarged, armed forelegs to grasp the *Koptothrips* and pierce their cuticle ($n = 33$ attacks in 8 h of observation: 30 by micropterous females, two by a micropterous male and one by a foundress). Attempted grasps were unsuccessful in 23 cases, but in 10 cases the *O. tepperi* held the *Koptothrips* in its forelegs; the *Koptothrips* escaped from the grasp in five of these cases, but on the other five occasions it was held and killed. *Koptothrips* also attempted to grasp *O. tepperi*, and did so successfully 10 times; each time, the grasped *O. tepperi* died within several minutes. The attacks by *O. tepperi* on *Koptothrips* appeared spontaneous, in that *O. tepperi* attacked *Koptothrips* without having been attacked themselves, and on six occasions such unsolicited attacks led to the attacking *O. tepperi* being killed.

Micropterous females of *O. habrus* attacked *Koptothrips* on 10 occasions in two hours of observation. These attacks led to successful holds of at least several seconds five times, and in one case a grasped *Koptothrips* was killed. As in *O. tepperi*, attacks appeared to be spontaneous. Indeed, *Koptothrips* put together with *O. habrus* did not attack them but continually attempted to escape from the half-gall.

To analyse experimentally the interaction of *Koptothrips* with *O. tepperi* and *O. habrus*, *Koptothrips* females were added to galls containing *Oncothrips* adults and the galls were sealed overnight. Similar galls, opened and sealed without the addition of *Koptothrips*, served as controls. In *O. tepperi*, all of the galls contained several micropterous adults (mean, 5.8; s.d., 2.6).

Some or all micropterous adults were found dead the next day in eight (67%) of 12 galls into which two female *Koptothrips* had been added, but in none of five galls without them ($\chi^2 = 6.3$, $P = 0.012$). The *O. tepperi* foundress was found dead in three (30%) of 10 galls to which *Koptothrips* had been added, and in none of five galls without them ($\chi^2 = 1.9$, $P = 0.17$; in two galls, the foundress had died in the field before collection). One or both *Koptothrips* were dead in nine (75%) of the 12 *O. tepperi* galls. In *O. habrus*, five of the galls to which one *Koptothrips* was added contained micropterous adults (mean, 5.3; s.d., 2.2) and six contained only the foundress. For galls containing at least one micropterous adult, some or all of them were dead in five (83%) of six galls to which a female *Koptothrips* had been added, but none were dead in five galls without *Koptothrips* ($\chi^2 = 7.64$, $P = 0.0057$). Similarly, the *O. habrus* foundress was dead in eight (73%) of 11 galls to which a female *Koptothrips* had been added, but in none of the 10 galls without *Koptothrips* ($\chi^2 = 11.75$, $P = 0.0006$). The *Koptothrips* was alive in all six *O. habrus* galls without micropterous adults, and in only two (40%)

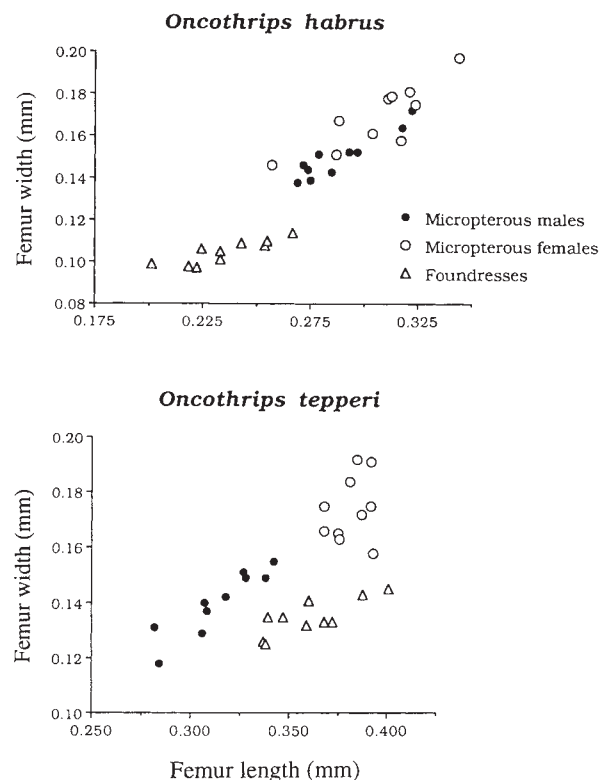


FIG. 1 Sexual and caste dimorphism in *Oncothrips tepperi* and *O. habrus*. At this life-history stage, galls of *O. habrus* ($n = 10$) contained 6–20 micropterous females (mean, 11.9; s.d. = 4.3); 0–5 micropterous males (mean 1.7, s.d. = 2.1); a single-winged foundress (in all galls); 0–10 micropterous pupae (mean 2.7, s.d. = 3.8); and 51–132 first and second instar larvae (mean 86.9, s.d. = 26.2). In *O. tepperi*, galls at this stage ($n = 10$) contained 3–6 micropterous females (mean 4.6, s.d. = 0.84); 0–2 micropterous males (mean 1, s.d. = 0.82); a single-winged foundress, 0–1 micropterous pupae (mean 0.5, s.d. = 0.53) and 57–98 larvae (mean 75.6, s.d. = 12.9). For each species, forelegs of adults from ten different galls were measured to the nearest 0.001 mm. Femur size was estimated as volume, $\pi r^2 h$, where r is femur width divided by two and h is femur length. In *O. habrus*, femur volume ($\text{mm}^3 \times 10^3$) varied significantly between foundresses (mean 2.05, s.d. = 0.39), micropterous females (mean 7.03, s.d. = 1.77) and micropterous males (mean 5.17, s.d. = 1.12; ANOVA, $F_{2,27} = 41.9$, $P < 0.001$; $P < 0.05$ for each pairwise comparison by Scheffé's test). In *O. tepperi*, femur volume also varied significantly between foundresses (mean 5.18, s.d. = 0.791), micropterous females (mean 9.13, s.d. = 1.33) and micropterous males (mean 4.91, s.d. = 1.07; ANOVA, $F_{2,27} = 47.1$, $P < 0.05$ comparing micropterous females with either foundresses or micropterous males; $P > 0.05$ comparing foundresses with micropterous males).

of the five galls with micropterous adults present ($\chi^2 = 4.95$, $P = 0.026$). These data demonstrate that *Koptothrips* kill *Oncothrips* and vice versa. A direct defensive role for micropterous adults of *O. habrus*, protecting the foundress, is suggested by the finding that the foundress was dead in all six of the galls without micropterous adults (where the *Koptothrips* was always alive), but she was dead in only two (40%) of the five galls with micropterous adults present ($\chi^2 = 4.95$, $P = 0.026$). Moreover, in these two galls with dead foundresses the micropterous adults were all dead ($n = 3$ in one gall, 8 in the other) or moribund ($n = 1$).

Do micropterous adults show a higher propensity than foundresses to attack invaders? Experimental removal of the first few adults that emerged from a hole in the gall, or attacked *Koptothrips* held at the hole, showed that in *O. habrus* micropterous females emerged from the gall and attacked *Koptothrips* with a higher probability than did foundresses (Table 1). By contrast, in *O. tepperi*, which exhibits less dimorphism in foreleg size than *O. habrus* (Fig. 1), foundresses and micropterous females did not differ significantly in tendency to emerge from the gall. In both species, micropterous males sometimes emerged or attacked invaders. Altruism by male haploids is not expected under inclusive fitness theory, unless the cost-to-benefit ratio for the selfish alternative is extremely high.

The observations and experiments described above indicate that micropterous adults of *O. tepperi* and *O. habrus* represent 'soldiers', defined as individuals morphologically or behaviourally specialized for defence. Soldiers are also found in aphids⁷, termites¹ and a polyembryonic wasp⁸, and in each of these taxa the soldiers are more or less sterile and the taxa are regarded as eusocial³. Do *O. tepperi* and *O. habrus* also exhibit eusociality?

Eusociality is normally defined by the presence of three traits: (1) overlap in generations between parents and their offspring; (2) cooperative brood care; and (3) reproductive division of labour, "with more or less sterile individuals working on behalf of fecund individuals"¹. In *O. tepperi* and *O. habrus*, the adult lives of the foundress and her micropterous offspring overlap substantially, and micropterous adults, by defending the gall and their gall-mates in an often self-sacrificing manner, are clearly engaging in cooperative brood care. Differences in reproduction between foundresses and micropterous females were analysed by dissecting females and ranking the size of their largest oocyte from one (undeveloped ovaries) to five (chorionated eggs). Many micropterous females had developing oocytes in both *O. tepperi* (67% of 273) and *O. habrus* (58% of 154), and some micropterous females contained fully developed chorionated eggs (12% of 273 in *O. tepperi*; 7% of 154 in *O. habrus*). But in both species micropterous females had significantly smaller oocytes on average than did foundresses (paired *t*-tests comparing foundresses with the mean for micropterous females, two samples for each species; $t = 2.94$, 3.81 , $P = 0.005$, 0.0012 , $n = 39$, 20 for *O. tepperi*; $t = 8.82$, 8.71 , $P < 0.0001$ for each, $n = 20$, 10 for *O. habrus*). Determining whether or not micropterous females have lower lifetime direct reproduction than foundresses would require analysis of maternity within complete adult broods. Similarity of reproductive rates is, however, highly unlikely because the offspring of the foundress alone, as micropterous adults and larvae, fill the gall almost completely: there is simply insufficient space in the gall for micropterous females to produce as many adult offspring as do foundresses.

The morphological specialization, defensive behaviour, and reduced reproduction of micropterous adults of *O. tepperi* and *O. habrus* indicate that these two species are eusocial. What selective pressures might have promoted eusociality among gall thrips in general, and in *O. tepperi* and *O. habrus* in particular? As in eusocial aphids⁷, naked mole-rats³, and many termites¹, gall thrips live and feed inside a highly valuable, persistent habitat that they have created, and this habitat is defensible primarily

by individuals specialized with weaponry and behaviour for heroic acts³. As in eusocial Hymenoptera, female gall thrips can facultatively bias the sex allocation ratio towards females^{4,9}, produce males by parthenogenesis⁹, and exhibit relatednesses to sisters of over one-half, all of which are presumed to favour the origin of eusocial behaviour^{10,11}. In *O. tepperi*, broods of macropterous adults are highly female-biased, and in both species, micropterous females in some galls are probably producing only sons as there are no micropterous males present with whom they could mate. Determining the importance of haplodiploidy in the origin and evolution of gall thrips eusociality requires further analysis of genetic relatedness and sex allocation ratios in eusocial and non-eusocial species. □

Received 29 June; accepted 8 September 1992.

1. Wilson, E. O. *The Insect Societies* (Belknap, Harvard Univ. Press, Cambridge, 1971).
2. Andersson, M. A. *Rev. ecol. Syst.* **15**, 165–189 (1984).
3. Alexander, R. D., Noonan, K. & Crespi, B. J. in *The Biology of the Naked Mole-Rat* (eds Sherman, P., Jarvis, J. U. M. & Alexander, R. D.) 3–44 (Princeton Univ. Press, New Jersey, 1991).
4. Crespi, B. J. *J. nat. Hist.* (in the press).
5. Mound, L. A. *Bull. Br. Mus. nat. Hist. (Ent.)* **25**, 389–466 (1971).
6. Mound, L. A. *Ent. Mon. Mag.* **105**, 159–162 (1970).
7. Aoki, S. in *Animal Societies: Theories and Facts* (eds Itô, Y., Brown, J. L. & Kikkawa, J.) 53–65 (Japan Sci. Soc. Tokyo, 1987).
8. Cruz, Y. P. *Nature* **294**, 446–447 (1981).
9. Crespi, B. J. in *Evolution and Diversity of Sex Ratio in Insects and Mites* (eds Wrensch, D. & Ebbert, M.) (Chapman and Hall, New York, in the press).
10. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
11. Trivers, R. L. & Hare, H. *Science* **191**, 249–263 (1976).

ACKNOWLEDGEMENTS. I thank the National Geographic Society for funding this research; R. H. Crozier and members of his laboratory for providing a friendly logistical base; E. Nielsen and I. Naumann of CSIRO Entomology for help with fieldwork; I. Clarke, G. C. Eickwort, C. D. Michener, L. A. Mound and L. Vawter for correspondence and discussion; and W. D. Hamilton, U. Mueller and P. W. Sherman for comments on the manuscript.

Wing bone stresses in free flying bats and the evolution of skeletal design for flight

Sharon M. Swartz*, Michael B. Bennett† & David R. Carrier*

* Program in Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA

† Department of Anatomical Sciences, The University of Queensland, Brisbane 4072, Australia

THE primary mechanical functions of limb bones are to resist deformation, and hence provide stiff levers against which muscles can act, and to be sufficiently strong to prevent breaking under static or dynamic loads which arise from normal and accidental activities¹. If bones perform these functions with a minimum amount of material, the energetic costs associated with building, maintaining and transporting the skeleton will be minimized². Appropriate skeletal architecture for minimizing mass while maximizing strength depends on forces imposed on structural elements. In the evolutionary acquisition of flight in the bat lineage, the forelimb skeleton must have come to experience locomotor-forces that differed from those engendered by the terrestrial locomotion of non-flying bat relatives. Here we successfully measure *in vivo* strain on the wing bones of flying mammals. Our data demonstrate that torsion and shear are unique and crucial features of skeletal biomechanics during flight, and suggest that the evolution of skeletal design in bats and other flying vertebrates may be driven by the need to resist these loads.

In typical mammalian locomotion, bones experience repetitive impacts in which forces are closely aligned to the limb's axis. In contrast, during flight, wing muscle contraction and pressure gradients applied by the wing membranes during the generation of lift and thrust apply forces to the long, slender

TABLE 1 Mean maximum peak principle strains magnitudes ($\mu\epsilon$) and their orientations

Humerus									
	Dorsal			Ventral			Medial		
	Magnitude	Orientation		Magnitude	Orientation		Magnitude	Orientation	
Mid-downstroke	1,554 (83)	21° (2)		1,496 (46)	22° (1)		2,004 (49)	21° (0)	
Bottom of downstroke	1,188 (37)	31° (1)		988 (32)	21° (3)		1,852 (45)	22° (2)	
Top of upstroke	488 (23)	0° (5)		572 (22)	2° (2)		468 (21)	25° (1)	
Radius									
	Dorsal			Dorsolateral			Ventral		
	Magnitude	Orientation		Magnitude	Orientation		Magnitude	Orientation	
Mid-downstroke	-1,808 (108)	28° (1)		-2,184 (47)	32° (1)		1,275 (44)	32° (1)	
Bottom of downstroke	-2,109 (95)	27° (1)		-1,912 (60)	32° (1)		1,316 (23)	29° (1)	
Top of upstroke	-405 (58)	29° (3)		904 (40)	21° (2)		243 (37)	-10° (3)	
Stress									
	Dorsal			Ventral			Medial		
	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear
Mid-downstroke	15.5 (0.9)	-6.8 (1.0)	4.2 (0.3)	14.4 (0.5)	-5.5 (0.2)	10.5 (0.5)	30.2 (1.0)	-3.0 (0.5)	15.7 (0.7)
Bottom of downstroke	-9.1 (0.8)	10.8 (0.9)	10.9 (0.8)	12.0 (0.4)	-4.1 (0.1)	5.2 (0.2)	28.3 (1.1)	6.2 (0.7)	10.7 (0.6)
Radius									
	Dorsal			Dorsolateral			Ventral		
	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear
Mid-downstroke	-36.1 (0.7)	6.8 (0.7)	-8.5 (0.7)	-45.8 (0.6)	-2.1 (0.1)	-11.2 (0.3)	20.6 (1.2)	-7.9 (0.7)	5.7 (0.2)
Bottom of downstroke	-37.8 (0.9)	2.2 (0.8)	-8.4 (0.8)	-36.6 (0.9)	-1.7 (0.1)	-7.7 (0.2)	17.7 (0.6)	-6.0 (0.8)	2.1 (0.4)

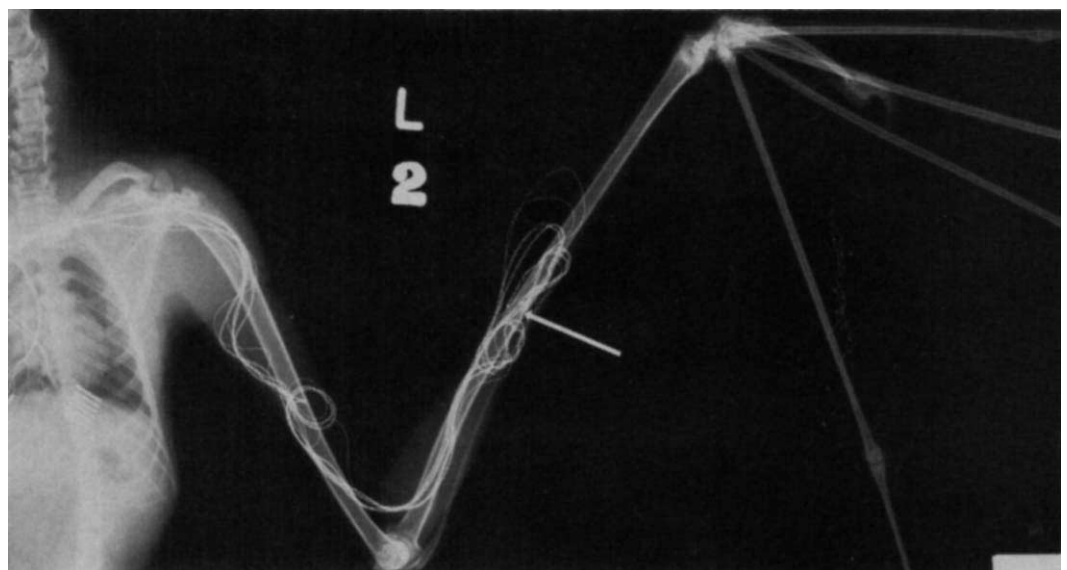
Each value represents the mean for 2–4 animals over 50 to 119 wingbeats; standard errors follow means in parentheses. Tensile strains are positive and compressive strains are negative. Mean magnitude of stresses (MPa) along the bones' primary material axes for the high stress (mid-downstroke and bottom of downstroke) portions of the wingbeat cycle. Calculations based on transversely orthotropic model for material characteristics of mammalian compact cortical bone⁴.

wing bones, and this complex force distribution changes dynamically throughout the wingbeat cycle³. Because of the many unknowns in the mechanical loading on the wings, it is impossible to make accurate and detailed predictions concerning the nature of strains experienced by wing bones during flight. To assess bone loading directly, we carried out the first successful study of natural patterns of bone deformation during free flight in large bats. We surgically implanted rosette strain gauges on the dorsal, ventral and medial or dorsolateral midshaft surfaces of the humerus and radius of four wild-caught grey-headed flying foxes (*Pteropus poliocephalus*; 500–850 g) (Fig. 1). We characterized bone loading throughout the wingbeat cycle by calculating principal strain magnitudes and their orientations

to each bone's longitudinal axis, and coordinated these recordings with simultaneous video records of the bats' movements. Bone stresses were then calculated using a transversely orthotropic model for compact bone mechanical properties⁴, with the assumption that the material characteristics of cortical bone of large bats is similar to that of birds and other mammals.

Strains recorded from each bone indicate complex loading regimes. Principal strain profiles for each recording site show a triphasic pattern, with peak loads occurring at the middle of the downstroke, the bottom of the downstroke and the top of the upstroke. Between these peaks, strain magnitudes drop considerably, approaching zero during the upstroke (Fig. 2). Peak strain magnitudes (Table 1) are somewhat lower than those

FIG. 1 Radiograph of grey-headed flying fox wing (L2 indicates left wing, subject number 2) implanted with strain gauges. We mist-netted all subjects at the Veterinary Science Farm of the University of Queensland, Brisbane, Australia; they were trained to fly the length of an outdoor flight cage (30 × 2 × 2 m) without stopping. Small rosette strain gauges (Tokyo Sokki Kenkyujo Co. FRA-1) were attached with a self-catalysing cyanoacrylate adhesive to the dorsal, ventral, and dorsolateral or medial midshaft cortices of the right humerus and left radius, as near as possible to midshaft but in areas free of direct muscle attachment to minimize localized distortions in overall bone strain due to muscle pull. The procedure was similar to that previously described^{9,18}. Gauge wires were subcutaneously passed proximally from the forearm or arm, ventrally in front of the shoulder, externalized through the skin superficial to the sternum, and soldered to connectors; sufficient slack was left to permit large amplitude wing movements without pulling gauges off the bones. After post-operative recovery, 6–10 channels of



strain data were amplified (MicroMeasurements Model 2120AK strain gauge conditioner/amplifiers) and sampled simultaneously at 250–400 Hz on a Macintosh IIfx computer (using National Instruments LabVIEW 2 data acquisition software and a National Instruments NB-MIO-16XL-18 A/D board) by means of a light-weight cable as each bat flew the length of the flight cage.

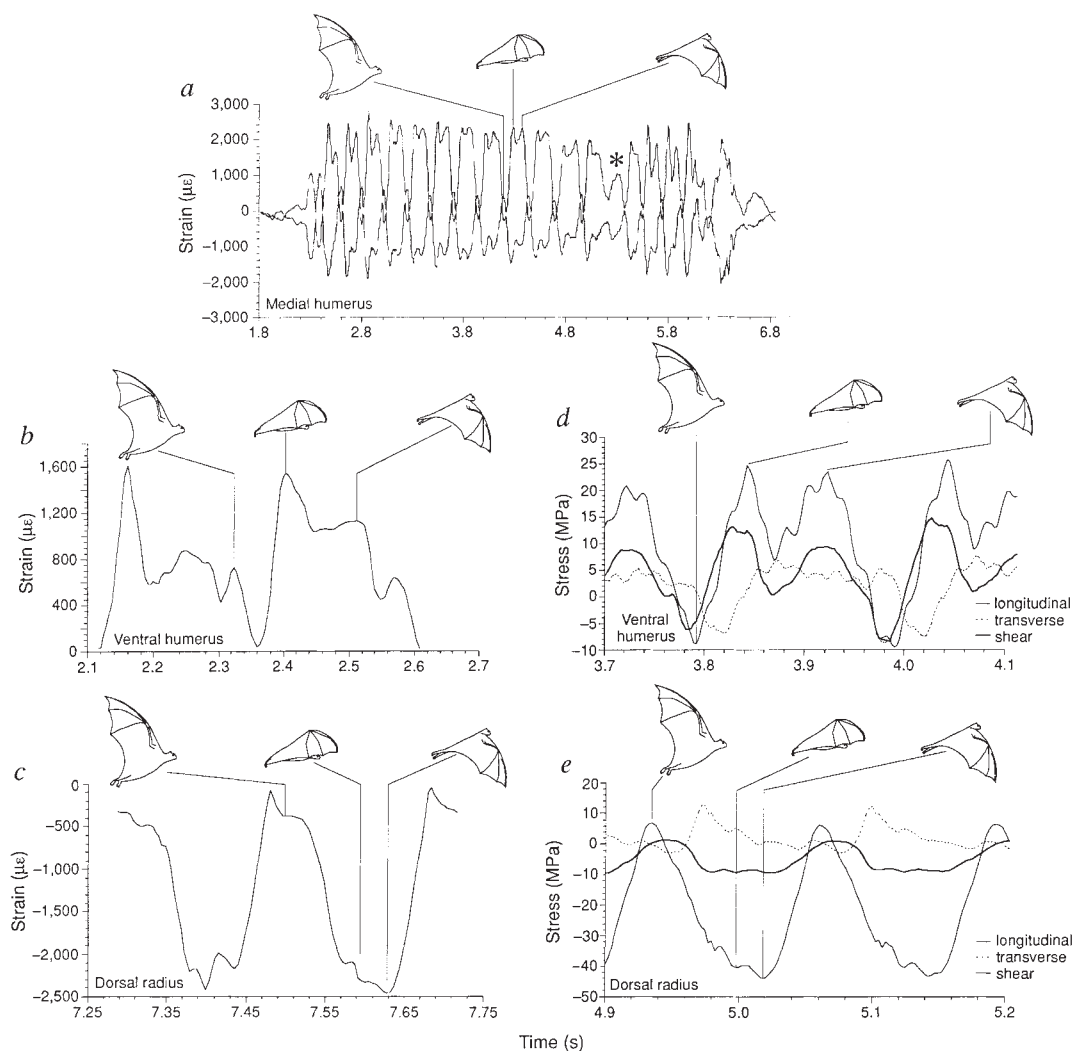
recorded from terrestrial mammals during vigorous locomotor activities (60–85% of previously observed values)^{5–7}. Peak principal strains at mid-downstroke and the bottom of downstroke are oriented at a substantial angle to the bone's long axis, with overall mean orientations of 23° for the humerus and 30° for the radius (orientation of 0° is aligned with the bone's axis, and positive deviations indicate the principal strain axis is rotated in a clockwise direction, with the axis extending along a proximomedial to distolateral direction with the limb in natural flight posture). In keeping with an earlier theoretical model of strength of wing bones in birds in relation to their function to resist aerodynamic forces during flight⁸, this pattern of off-axis strain of similar magnitude around the bone circumference is suggestive of significant torsional loading. Similar torsional loading has recently been observed in studies of the pigeon humerus during flapping flight (A. A. Biewener and K. P. Dial, personal communication).

Longitudinal stresses calculated from the raw strain data (Table 1; Fig. 3) are comparable in magnitude to stresses developed in other mammals during fast canters, gallops and some jumps^{9–11}. Transverse stresses are large, particularly at the bottom of the downstroke, reaching in the humerus and radius

respectively, maxima of 47% and 38% of the longitudinal value during the mid-downstroke and 119% and 34% at the bottom of the downstroke. Shear stresses are also significantly greater in bats than in other mammals (more than 15 MPa, compared with values of no greater than 6 MPa recorded during walking and running in human and dogs^{4,12}). These large transverse and shear stresses are both consistent with substantial torsional loads. These results differ significantly from the typical pattern of mammalian limb bone loading, in which bending predominates and torsion is of relatively little significance¹³. This is notable given that bone fails at far lower stress in shear (corresponding to about 65 MPa) than in tension or compression (corresponding to about 150 and 200 MPa respectively)¹⁴.

The mechanism that is most likely to be responsible for producing these torsional stresses in bat wing bones is probably one that is common to all vertebrates capable of active, powered flight. Because the skeleton is near the wing's leading edge in bats, as in all flying vertebrates, the wing's centre of pressure, at which the aerodynamic forces generated during flight act, lies posterior to the major bony supports. Upward directed forces at the centre of pressure thus exert a torsional moment in the pronating direction, transmitted through the wing membrane

FIG. 2 *a*, Representative maximum and minimum principal strain records from the medial humerus; *b*, maximum principal strain records from ventral humerus and *c*, dorsal radius, and *d*, calculated longitudinal, transverse and shear stresses from the ventral humerus and *e*, dorsal radius of one bat for single flights (60 frames per second) taken during recordings. Strain magnitudes are reported in microstrain (change in length/original length $\times 10^{-6}$); tensile (elongating) strains are positive and compressive (shortening) strains are negative. Stress values in megapascals (MPa) are calculated from raw strain data assuming that material characteristics of bat bone are comparable to those of other mammals and using a transversely orthotropic model for bone stiffness⁴. In *a*, a complete flight sequence is depicted from takeoff (small peaks at the far left), through a number of regular wingbeat cycles, through a postural shift in preparation for landing (asterisk), followed by 5 decelerating wingbeats and landing. In *b–e*, maximum principal strains and stresses in the bone's material axes are shown for two representative wingbeats. In *b*, *c* and *e*, magnitude changes continually throughout the locomotor cycle for each recording site; representative mid-downstroke, bottom of downstroke and top of upstroke phases of one wingbeat are indicated by the schematics. Bone strains typically reach their greatest absolute values at mid-downstroke and return



to smaller peak values at the bottom of the downstroke and again at the top of the upstroke. Strain values return close to zero during some phases of the locomotor cycle. Peaks in longitudinal stress correspond to those in principal strain; transverse and shear stress peaks may occur at somewhat different points in the wingbeat cycle.

and the extended digits, particularly digits 4 and 5 (ref. 8). In the case of birds, wing feathers may serve the same stress transmission role in a somewhat different fashion. A similar mechanism may be responsible for large transverse stresses, as the stresses in the taut wing are transmitted by means of the attachments of the wing membrane along each bone diaphysis.

The cross-sectional geometry of bat wing bones can be related directly to the need to resist torsional stresses imposed by propulsive downstroke movements. The ability of a beam-like structure to resist torsion depends on the magnitude of torque applied (T), the moment arm of the torsional force (r), and the polar second moment of area of the beam cross-section (J) such that the torsion stress (τ) is $\tau = Tr/J$, where J is $J = \pi(R^4 - r^4)/2$. For a given cross-sectional area, J is maximized by circular cross-sectional geometry of maximum outer diameter and minimum wall thickness.

Bats ought, therefore, to have wider, thinner-walled forelimb bones than terrestrial mammals. Our measurements indicate that the wing bones from which we recorded strains indicative of torsion are indeed thinner-walled than those of other mammals (Fig. 3). It has been previously noted that birds and pterosaurs have thin-walled wing bones¹⁵, with pterosaurs displaying especially marked cortical reduction in the metacarpals.

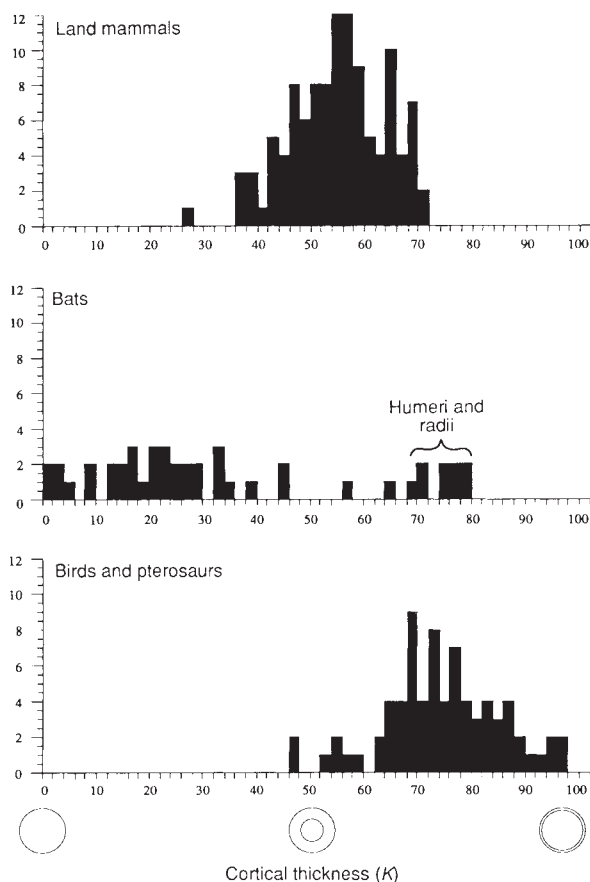


FIG. 3 Frequency distribution of cortical bone thickness for a variety of vertebrates (terrestrial mammal, bird and pterosaur data from ref. 12, additional bat data collected from radiographs of this study's subjects). Cortical thickness is expressed as K , and if R is the radius of the bone cross-section, then the marrow cavity thickness equals KR (after ref. 12). When K is zero the bone cross-section is solid, and K approaches 1 as wall thickness approaches zero. Distributions for land mammals and birds plus pterosaurs show some overlap but are clearly distinct, with larger mean K values for flying vertebrates. Bats are clearly distinct from other mammals; the humeri and radii investigated here show reduced thickness compared with other mammals and fall well within the bird/pterosaur range. Other bat wing bones (metacarpals and phalanges) also deviate from the typical mammalian pattern, with greatly increased cortical thickness.

Although this geometrical specialization is typically attributed to the need to reduce skeletal mass in flying animals¹⁶, the skeleton makes up a comparable proportion of total body mass in birds and mammals¹⁷, indicating that other factors are probably involved. We propose that the 'hollow' geometry of bird and pterosaur bones relates directly to an adaptive maximization of J to minimize torsional stress transmitted through wing membranes. Bats have converged upon the bird and pterosaur pattern, with cortical thickness values falling well within the range of birds and pterosaurs and only barely overlapping other mammals. Hence, reduction in thickness in cortical bone in vertebrate wings has evolved at least three times independently, and this fundamental aspect of skeletal design in limbs of flying vertebrates should be interpreted not only as a mechanism to reduce weight but, at least as importantly, as an adaptation to resist large torsional stresses. □

Received 20 July; accepted 8 September 1992.

- Alexander, R. M. Ker, R. F. & Bennett, M. B. *J. Zool., Lond.* **222**, 471-478 (1990).
- Alexander, R. M. *Sci. Prog., Oxf.* **67**, 109-130 (1981).
- Rayner, J. M. V. in *Recent Advances in the Study of Bats* (eds Fenton, M. B., Racey, P. & Rayner, J. M. V.) 23-42 (Cambridge Univ. Press, UK, 1987).
- Carter, D. R. *J. Biomech.* **11**, 199-202 (1978).
- Biewener, A. A. & Taylor, C. R. *J. exp. Biol.* **123**, 383-400 (1986).
- Biewener, A. A., Swartz, S. M. & Bertram, J. E. A. *Calc. Tiss. Int.* **39**, 390-395 (1986).
- Rubin, C. T. & Lanyon, L. E. *J. theor. Biol.* **107**, 321-327 (1984).
- Pennycuik, C. J. *J. exp. Biol.* **46**, 219-233 (1967).
- Biewener, A. A., Thomason, J. & Lanyon, L. E. *J. Zool., Lond.* **201**, 67-82 (1983).
- Biewener, A. A., Thomason, J. & Lanyon, L. E. *J. Zool., Lond.* **214**, 547-565 (1988).
- Rubin, C. T. & Lanyon, L. E. *J. exp. Biol.* **101**, 187-212 (1982).
- Bertram, J. E. A. & Biewener, A. A. *J. theor. Biol.* **131**, 75-92 (1988).
- Nordin, M. & Frankel, V. H. in *Basic Biomechanics of the Musculoskeletal System* (eds Nordin, M. & Frankel, V. H.) 3-29 (Lea & Febiger, Philadelphia and London, 1989).
- Currey, J. D. & Alexander, R. M. *J. Zool. Lond.* **206**, 453-468 (1985).
- Pough, F. H., Heiser, J. B. & McFarland, W. N. *Vertebrate Life* (Macmillan, New York, 1985).
- Prange, H. D., Anderson, J. F. & Rahn, H. *Am. Nat.* **113**, 103-122 (1979).
- Lanyon, L. E. *Acta orthop. belg.* **42**, 98-108 (1976).

Positive selection of T-lymphocytes induced by intrathymic injection of a thymic epithelial cell line

Stanislav Vukmanović*, Andres G. Grandea III*, Susan J. Faas†, Barbara B. Knowles† & Michael J. Bevan*

* Department of Immunology and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

† The Wistar Institute of Anatomy and Biology and University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

T LYMPHOCYTES recognize antigens as peptide fragments associated with molecules encoded by the major histocompatibility complex (MHC) and expressed on the surface of antigen-presenting cells¹. In the thymus, T cells bearing $\alpha\beta$ receptors that react with the MHC molecules expressed by radioresistant stromal elements are positively selected for maturation²⁻⁵. In ($A \times B \rightarrow A$) bone marrow chimaeras, T cells restricted to the MHC-A haplotype are positively selected, whereas MHC-B-reactive thymocytes are not. We investigated whether the introduction of particular thymic stromal elements bearing MHC-B molecules could alter the fate of B-reactive T cells in these ($A \times B \rightarrow A$) chimaeras. Thymic epithelial cell (TEC) lines expressing H-2^b were introduced by intrathymic injection into ($H-2^{b/s} \rightarrow H-2^s$) bone marrow chimaeras and we measured their ability to generate H-2^b-restricted cytotoxic T-lymphocytes (CTLs). We report here that one TEC line, 427.1, was able positively to select CTLs specific for influenza and vesicular stomatitis virus antigens in association with class I H-2^b molecules. In addition, line 427.1 can process cytoplasmic proteins

for presentation to H-2K^b- and H-2D^b-restricted CTLs. Thus, a TEC line capable of normal class I MHC antigen processing and presentation *in vitro* can induce positive selection after intrathymic injection.

TEC lines 427.1 and 1308.1 were derived from Simian Virus 40 (SV40) T antigen transgenic (B6×SJL)_F₂ mice (ref. 6 and S.J.F., J. L. Rothstein and B.B.K., manuscript in preparation). Both cell lines express H-2K^b and H-2D^b class I MHC molecules (data not shown). The ability of these cell lines positively to select H-2^b-restricted CTLs was tested by intrathymic injection into ((B6×SJL)_F₁→SJL) chimaeras 11 days after bone marrow reconstitution. Some weeks after intrathymic injection, the chimaeras were immunized with influenza virus and, after *in vitro* restimulation with infected B6 spleen cells, H-2^b-restricted CTL responses were measured. In (F₁→F₁) chimaeras a strong influenza-specific cytotoxic response was observed (Fig. 1a), whereas (F₁→SJL) chimaeras were unresponsive (Fig. 1b). (F₁→SJL) chimaeras injected intrathymically with 1308.1 TECs were similarly unresponsive (Fig. 1c). In contrast, T cells from (F₁→SJL) chimaeras injected intrathymically with 427.1 TECs generated a strong CTL response to influenza-infected H-2^b targets (Fig. 1d). The MHC restriction and peptide specificity of the anti-influenza response is shown in Table 1, experiment 1. In H-2^b mice, influenza-specific CTLs recognize the nucleoprotein-derived nonamer peptide 366–374 presented by H-2D^b (refs 7, 8). Cytotoxic effectors generated both from F₁ mice and from (F₁→SJL) chimaeras injected intrathymically with 427.1 TECs were able to kill influenza-infected H-2^b targets but not influenza-infected H-2^d or H-2^s targets. Furthermore, H-2^b target cells were killed

if the synthetic peptide corresponding to nucleoprotein 366–374 was added to the CTL assay.

To investigate whether the TEC line 427.1 can positively select T cells restricted by another class I molecule, chimaeras were immunized with vesicular stomatitis virus (VSV). The CTL response to VSV is directed to the nucleocapsid (N) protein 52–59 peptide presented by H-2K^b (ref. 9). F₁ mice and (F₁→SJL) chimaeras injected intrathymically with 427.1 cells, but not uninjected (F₁→SJL) chimaeras, could kill VSV-infected H-2^b cells, but not VSV-infected H-2^d or H-2^s targets (Table 1, experiment 2). The addition of VSV N 52–59 synthetic peptide, but not chicken ovalbumin (OVA) 257–264, to the CTL assay rendered H-2^b targets sensitive to these anti-VSV effectors.

To assay for the presence of TECs in the thymus of injected mice, thymic DNA was isolated from mice immunized with influenza or VSV. A 0.5-kilobase (kb) fragment of the sequence encoding SV40 T antigen¹⁰ was amplified using polymerase chain reaction (PCR) from the thymus of chimaeras injected with line 427.1 (lane 5, Fig. 2a) but not from the thymus of (B6×SJL)_F₁ mice, uninjected or 1308.1-injected (F₁→SJL) chimaeras (lanes 3, 4 and 6, respectively). To assess the sensitivity of detection, decreasing amounts of DNA isolated from line 427.1 were mixed with DNA isolated from the normal (B6×SJL)_F₁ thymus, and SV40 T antigen sequences were amplified (Fig. 2b). The amount of DNA isolated from the equivalent of 800 427.1 cells provided sufficient template to generate visible PCR product (lane 4). Because we have routinely analysed 1% of the DNA from the thymus of injected mice, we can estimate that 8×10⁴ cells was the minimum number of 427.1 cells present in the thymus (8% of the injected number). Therefore, TEC line

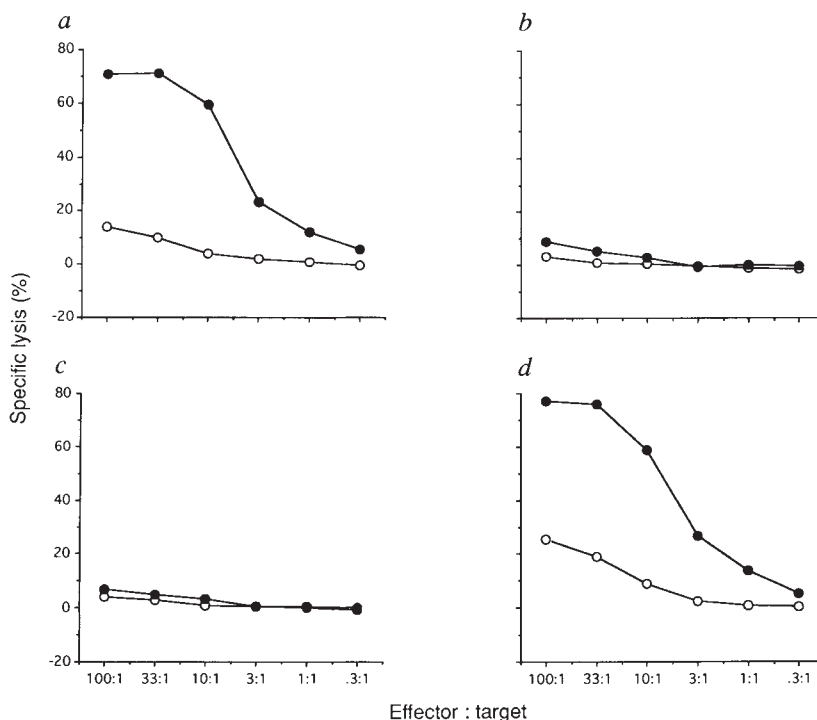


FIG. 1 The effect of intrathymic injection of H-2^b-expressing TEC lines into (F₁→SJL) bone marrow chimaeras on positive selection of CTLs recognizing influenza virus plus H-2^b. (B6×SJL/J)_F₁ (a) or SJL/J (b–d) female mice were lethally irradiated (850 rads), and reconstituted with intravenous injection of 5×10⁶ (B6×SJL/J)_F₁ T-cell depleted bone marrow cells. After reconstitution (11 days) mice were left untreated (a, b) or injected intrathymically with 1×10⁶ 1308.1 (c) or 427.1 (d) epithelial cells. Five weeks after this the mice were immunized by intraperitoneal injection of 250 haemagglutinating units (HAU) influenza A/PR8/34 virus. Ten days after immunization spleen cells were restimulated *in vitro* for 5 days with irradiated (2,000 rads) influenza-infected B6 spleen cells. Cytotoxic activity against untreated (open

symbols) or influenza-infected (filled symbols) EL4 (H-2^b), ⁵¹Cr-labelled target cells was tested in a standard 4-h ⁵¹Cr-release assay.

METHODS. The isolation, and characterization of TEC lines 427.1 and 1308.1, that express markers found on subcapsular and cortical epithelium, respectively, will be described elsewhere (S.J.F., J. L. Rothstein and B.B.K., manuscript in preparation). Bone marrow cells were depleted of Thy-1⁺ cells by T24 antibody²⁹ and complement treatment. Intrathymic injections were done as in ref. 30. Some 5×10⁵ TECs were injected per lobe. For the *in vitro* stimulation, 1×10⁸ B6 spleen cells were infected with 1×10³ HAU A/PR8/34. For CTL assays, 1×10⁶ EL4 cells were simultaneously infected with 1×10² HAU A/PR8/34 and labelled with ⁵¹Cr.

FIG. 2 Detection of TEC-derived SV40 T antigen DNA in the thymi of injected mice. *a*, Chimaeras were made, injected with TECs, and immunized as described in Fig. 1. Thymi were taken from mice killed 9–13 days after immunization. DNA isolated from TEC line 427.1 (lane 1) or 1308.1 (lane 2), and thymic DNA from (B6 × SJL)_F₁ (lane 3), (F₁ → SJL) chimaera (lane 4), (F₁ → SJL) chimaera injected with 427.1 cells intrathymically (lane 5), (F₁ → SJL) chimaera injected with 1308.1 cells intrathymically (lane 6), was subjected to PCR using primers derived from SV40 T antigen sequence. Using primers derived from the genomic sequence of H-2K^b MHC class I molecule, a 0.5-kb fragment could be amplified from all samples (data not shown). *b*, DNA was isolated from line 427.1 and F₁(B6 × SJL/J) thymus. DNA equivalents of 10⁵ (lane 1), 2 × 10⁴ (lane 2), 4 × 10³ (lane 3), 8 × 10² (lane 4), or 1.6 × 10² (lane 5), or 0 427.1 cells (lane 6) were mixed with thymus DNA. DNA (2.5 µg) in each reaction was subjected to PCR reaction using SV40 T antigen primers.

METHODS. Thymi were homogenized and lysed with 0.5% SDS, and DNA was isolated by phenol-chloroform extraction of Proteinase K/RNase A-treated lysates. PCR was done in 100 µl 10 mM Tris, pH 8.3, 50 mM KCl, 2 mM MgCl₂. Primers (5'-GTGCCCTTACATCCTC-3' and 5'-CAGCCAC-TATAAGTACC-3' derived from the SV40 T antigen sequence¹⁰) were used at 20 µg ml⁻¹, each dNTP at 240 µM, *Taq* DNA polymerase at 25 U ml⁻¹. DNA (2–2.5 µg) was used as template. Samples were subjected to 30 cycles of amplification, each consisting of denaturation at 94 °C for 60 s, annealing at 48 °C for 120 s, and polymerization at 72 °C for 90 s. One tenth of the reaction was analysed by 1% agarose gel electrophoresis.

427.1 had the capacity to remain in the thymus weeks after injection, whereas line 1308.1 was not detectable. In other experiments, two other TEC lines, as well as macrophage cell line IC-21 also failed to remain detectable in the thymus (data not shown).

It has been suggested that the MHC molecules on the cells that induce positive and negative selection express different sets of peptides, so that positive and negative selection depend on different ligands for the TCR^{11–14}. Some mature T cells reactive with self-MHC plus a foreign peptide also recognize unmodified syngeneic thymic epithelium¹². Furthermore, a peptide-MHC complex expressed abundantly in the periphery is hardly detectable in thymus cortical epithelium¹³. The peptides presented by MHC molecules influence positive selection^{15–17}. In the case of MHC class I, the range of peptides bound to a given class I molecule can be substantially different dependent on whether the MHC-encoded transporter molecules are operative^{18–20}. In cell lines with defects in the peptide transporter, only a few peptides can enter the endoplasmic reticulum lumen and associate with class I^{18,19}, and thus the presentation of cytoplasmic antigens is less efficient^{21–23}. We investigated whether there was

a defect in normal class I presentation in the positively selecting cell line 427.1. Cells were infected with influenza virus or VSV and tested for sensitivity to the appropriate CTL lines. Also, OVA was introduced into the cytoplasm of 427.1 cells by osmotic loading²⁴ and assayed for recognition by OVA-specific H-2K^b-restricted CTLs. All three effectors could lyse infected/loaded 427.1 targets (Fig. 3*a–c*). In addition, Brefeldin A, an agent that blocks the normal class I antigen-processing pathway^{25,26}, completely abrogated antigen presentation of the three endogenous antigens by 427.1 cells, while allowing the presentation of exogenously added synthetic peptides (Fig. 3*a–c*).

To test whether self-peptides associated with class I molecules in other tissues are also present in line 427.1, peptide-specific alloreactive CTL clones were used. These clones are specific for H-2K^b binding peptides from spleen recognized in the context of a mutant H-2K^b molecule (Gln 65 → Asp 65), designated D65 (ref. 27). Targeting peptides were contained in reversed-phase high-performance liquid chromatography (HPLC) fractions 32 (for clone d3.1) and 30 and 33 (for clone d3.3) in acid extracts of H-2K^b isolated from B6 spleen cells²⁷. Immunoprecipitated

TABLE 1 Intrathymic injection of the H-2^{b/s} TEC line 427.1 into (F₁ → SJL) bone marrow chimaeras (H-2^{b/s} → H-2^s) induces positive selection of H-2^b-restricted, influenza and VSV-specific cytotoxic T-cells

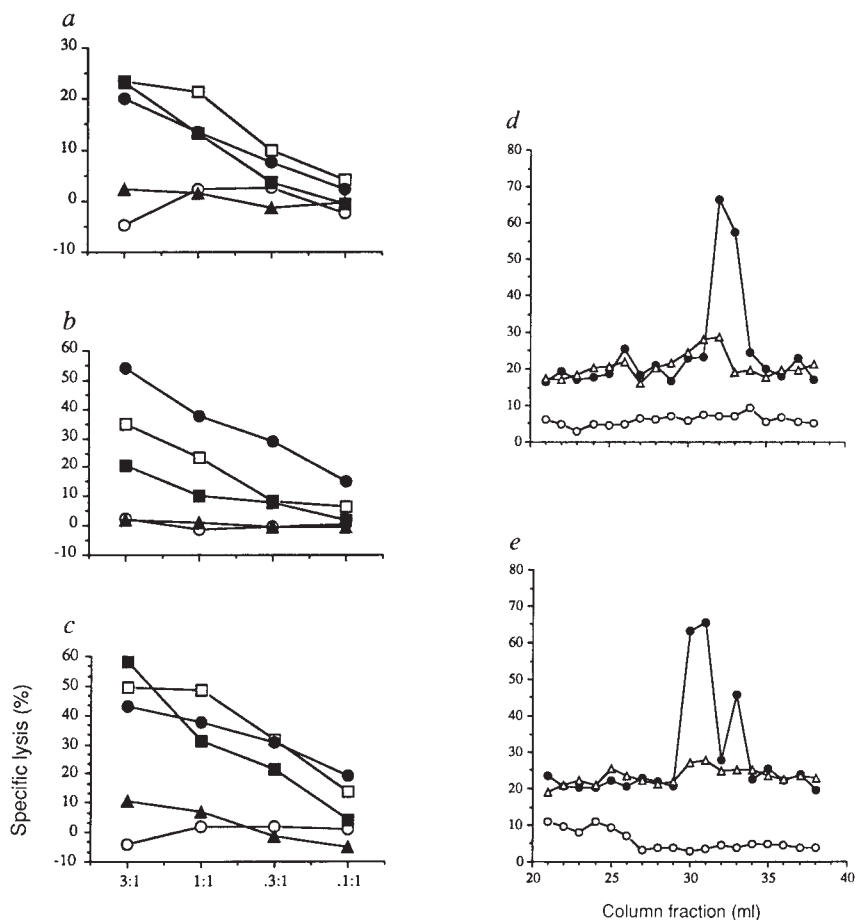
				Specific lysis (%)								
Experiment number	Chimaera	Cell line injected i.t.	Immunization	—	EL4		OVA 257–264	—	P815		B11	P815*
					A/PR8/34-infected	PR8 NP 366–374			A/PR8/34-infected	—	A/PR8/34-infected	
1	(B6 × SJL/J)F ₁	None	A/PR/8/34	22/7	66/45	48/32	0/–2	6/3	8/1	19/4	22/7	84/23
	F ₁ → SJL/J	None	A/PR/8/34	4/1	8/2	7/0	8/–1	5/1	8/1	7/1	7/5	64/14
	F ₁ → SJL/J	427.1	A/PR/8/34	6/1	44/14	24/9	–2/–2	13/3	12/3	12/7	16/10	79/20
2	(B6 × SJL/J)F ₁	None	VSV	—	VSV-infected	VSV N 52–59	OVA 257–264	—	VSV-infected	—	VSV-infected	
		10/2		31/20	28/20	6/1	13/3	10/3	23/11	17/8	79/43	
		6/2		7/3	10/10	10/1	5/1	3/3	7/4	13/5	74/42	
		7/1		27/11	27/12	5/0	7/3	5/2	14/2	10/9	68/34	

Chimaeras were made as in Fig. 1 legend. Three to four weeks after TEC-injection mice were immunized by intraperitoneal injection with 250 HAU influenza A/PR8/34 virus or 5 × 10⁶ plaque-forming units VSV. After immunization (9–13 days) spleen cells were restimulated *in vitro* for 5 days with irradiated (2,000 rads) influenza- or VSV-infected B6 spleen cells. Cytotoxic activity against untreated, influenza- or VSV-infected EL4 (H-2^b), P815 (H-2^d) or B11 (H-2^s) ⁵¹Cr-labelled target cells was tested in a 4 h ⁵¹Cr-release assay. Synthetic peptides corresponding to A/PR8/34 nucleoprotein 366–374⁸, VSV N 52–59⁹, or chicken ovalbumin (OVA) 257–264³¹ sequences were added to the CTL assay at 50 nM. Per cent specific lysis at effector:target ratios of 50:1/5:1 are shown. i.t., intrathymic; NP, nucleoprotein.

* Spleen cells were stimulated *in vitro* with irradiated allogeneic DBA/2 (H-2^d) spleen cells.

FIG. 3 TEC line 427.1 can process antigens for presentation to class I MHC-restricted CTLs. *a-c*, Influenza virus-specific CTL line (*a*), VSV-specific CTL line (*b*), or ovalbumin-specific CTL clone GA4 (*c*) were tested for the ability to lyse 427.1 cells in the absence (open circles) or presence (open squares) of the appropriate synthetic peptide, and 427.1 cells infected with influenza (*a*), VSV (*b*) or loaded with OVA (*c*) (filled circles). Infected/loaded 427.1 cells were treated with Brefeldin A in the absence (filled triangles) or presence (filled squares) of the appropriate synthetic peptide. In preliminary experiments, 427.1 cells showed slightly but consistently lower targeting activity compared with infected/loaded EL4 cells (data not shown). *d, e*, Reversed-phase HPLC-separated fractions from anti-H-2K^b (circles) and anti-H-2D^b (triangles) immunoprecipitates of 427.1 cell lysates were analysed for targeting activity for CTL clones d3.1 (*d*) and d3.3 (*e*) using ⁵¹Cr-RMA-S (open circles) or D65-transfected RMA-S (filled circles and open triangles) as targets.

METHODS. The anti-Hk/68 CTL line was established by priming F₁(B6×SJL) mice with 250 HAU of influenza Hk/68 virus. Two weeks later, spleen cells were restimulated with Hk/68-infected, irradiated (2,000 rads) B6 spleen cells. The line was maintained by weekly restimulation with Hk/68-infected, irradiated B6 spleen cells in the presence of 10% rat concanavalin A supernatant. The establishment and maintenance of the anti-VSV CTL line and the CTL clone GA4 were described previously^{22,32}. Cells (1×10⁶) from line 427.1 were infected by incubation in the presence of 100 HAU influenza Hk/68 or 250×10⁶ p.f.u. VSV. Osmotic loading of OVA was done as described²⁴. Cells were labelled with ⁵¹Cr during infection, or after osmotic loading. Brefeldin A was used at 10 µg ml⁻¹ during infection and/or labelling with ⁵¹Cr, and at 2.5 µg ml⁻¹ during CTL assay. Synthetic peptides from influenza nucleoprotein, VSV N or OVA, as in Table 1, were added to the CTL assay at 50 nM. To make peptide extracts of class I molecules, 1×10⁸ 427.1 cells were lysed and the H-2K^b and H-2D^b were immunoprecipitated as described²⁷. Acid extraction of the washed



precipitates, fractionation of the low molecular mass material on C18 reversed phase HPLC, and the targeting assay of the fractions has been described²⁷. Derivation of the RMA-S cell line expressing the D65 mutant H-2K^b molecule and the CTL clones d3.1 and d3.3 has been described²⁷.

H-2K^b molecules from line 427.1 also yielded material with the same HPLC elution profile able to sensitize D65 transfected RMA-S cells for recognition by these CTL clones (Fig. 3*d, e*). Thus, the 427.1 cell line seems to have a similar representation of self-peptides in its MHC class I grooves, compared to peripheral cells.

We have shown that the TEC line 427.1, which can 'conventionally' process and present endogenous antigens to CTLs *in vitro*, can induce positive selection of H-2^b-restricted CTLs *in vivo* when injected intrathymically. One implication of this is

that the MHC-peptide ligands involved in positive and negative selection are the same. If this is the case, the range of T-cell receptors that are positively selected is broader than that deleted during negative selection^{11,28}. These findings do not necessarily imply that all class I MHC-associated peptides of this cell line are same as those on bone marrow-derived cells, there are probably tissue-specific peptides expressed. Using 427.1 cells with controlled expression of genes it may be possible to investigate further the requirements for positive selection. □

Received 11 June; accepted 14 August 1992.

1. Yewdell, J. W. & Bennink, J. R. *Adv. Immun.* (in the press).
2. Bevan, M. J. *Nature* **269**, 417-418 (1977).
3. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882-896 (1978).
4. Sprent, J. *J. exp. Med.* **146**, 1838-1842 (1978).
5. Kisielow, P., Teh, H. S., Bluthmann, H. & von Boehmer, H. *Nature* **335**, 730-733 (1988).
6. Brinster, R. L. *et al. Cell* **37**, 367-379 (1984).
7. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
8. Rotzschke, O. *et al. Nature* **348**, 252-254 (1990).
9. Van Bleek, G. M. & Nathenson, S. G. *Nature* **348**, 213-216 (1990).
10. Fiers, W. *et al. Nature* **273**, 113-120 (1978).
11. Marrack, P. & Kappler, J. *Immun. Today* **9**, 308-315 (1988).
12. Marrack, P., McCormack, J. & Kappler, J. *Nature* **338**, 503-505 (1989).
13. Murphy, D. B. *et al. Nature* **338**, 765-768 (1989).
14. Mizuuchi, T., Kasai, M., Kokuho, T., Kakiuchi, T. & Hirokawa, K. *J. exp. Med.* **175**, 1601-1605 (1992).
15. Nikolic-Zugic, J. & Bevan, M. J. *Nature* **344**, 65-67 (1990).
16. Sha, W. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6186-6190 (1990).
17. Berg, L. J., Frank, G. D. & Davis, M. M. *Cell* **60**, 1043-1053 (1990).
18. Henderson, R. A. *et al. Science* **255**, 1264-1266 (1992).

19. Wei, M. & Cresswell, P. *Nature* **356**, 443-446 (1992).
20. Powis, S. J. *et al. Nature* **357**, 211-215 (1992).
21. Townsend, A. R. M. *et al. Nature* **340**, 443-448 (1989).
22. Hosken, N. A. & Bevan, M. J. *Science* **248**, 367-370 (1990).
23. Spies, T. *et al. Nature* **355**, 644-646 (1992).
24. Moore, M. W., Carbone, F. R. & Bevan, M. J. *Cell* **54**, 777-785 (1988).
25. Nuchtern, J. G., Bonifacio, J. S., Biddison, W. E. & Klausner, R. D. *Nature* **339**, 223-226 (1989).
26. Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072-1075 (1989).
27. Grandea, A. G. & Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2794-2798 (1992).
28. Fink, P. *Immun. Today* **9**, 377-380 (1988).
29. Dennert, G., Hyman, R., Lesley, J. & Trowbridge, I. *Cell. Immun.* **53**, 350-364 (1980).
30. Goldschneider, I., Komschlies, K. L. & Greiner, D. L. *J. exp. Med.* **163**, 1-17 (1986).
31. Rotzschke, O. *et al. Eur. J. Immun.* **21**, 2891-2894 (1991).
32. Nikolic-Zugic, J. & Carbone, F. R. *Eur. J. Immun.* **20**, 2431-2437 (1990).

ACKNOWLEDGEMENTS. We thank J. Nikolic-Zugic and P. Fink for advice on intrathymic injections, J. Thorbecke for the B11 cell line, D. Debelak for technical assistance and J. Harty, S. Jameson, K. Hogquist and E. Pamer for discussion and comments. Supported by the NIAID and the HMI.

Quantal transmitter secretion from myocytes loaded with acetylcholine

Yang Dan & Mu-ming Poo*

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

* To whom correspondence should be addressed

IT is well known that transmitter secretion requires specialized secretory organelles, the synaptic vesicles, for the packaging, storage and exocytotic release of the transmitter^{1,2}. Here we report that when acetylcholine (ACh) is loaded into an isolated *Xenopus* myocyte, there is spontaneous quantal release of ACh from the myocyte which results in activation of its own surface ACh channels and the appearance of membrane currents resembling miniature endplate currents. This myocyte secretion probably reflects Ca^{2+} -regulated exocytosis of ACh-filled cytoplasmic compartments. Furthermore, step depolarization of the myocyte membrane triggers evoked ACh release from the myocyte with a weak excitation-secretion coupling. These findings suggest that quantal transmitter secretion does not require secretory pathways unique to neurons and that the essence of presynaptic differentiation may reside in the provision of transmitter supply and modification of the pre-existing secretion pathway.

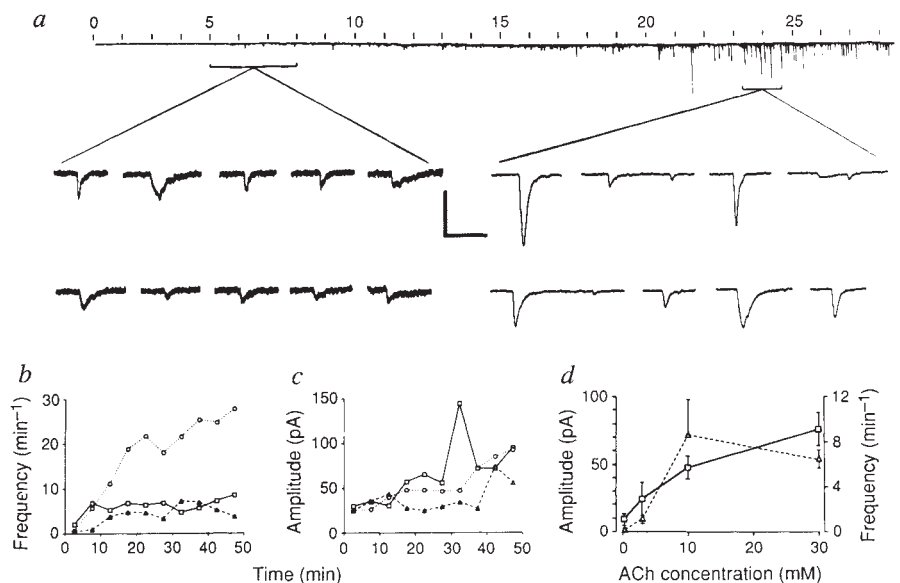
Experiments were carried out on spherical *Xenopus* myocytes, which express high levels of ACh receptors on their surfaces one day in culture^{3,4}. A whole-cell recording method⁵⁻⁷ was used to load ACh into isolated myocytes and to record the myocyte membrane current. With 10 mM ACh added in the pipette solution, inward pulsatile currents resembling miniature

endplate currents (m.e.p.cs) appeared within a few minutes, with a gradual increase in the frequency and amplitude (Fig. 1a-c). The current events showed a dose-dependence on the concentration of ACh in the pipette in the range 0.3 mM to 30 mM (Fig. 1d), reaching the estimated free ACh levels within cholinergic nerve terminals⁸⁻¹¹. The distribution of the current amplitudes was skewed towards smaller sizes, similar to that of the m.e.p.cs observed at innervated myocytes in these cultures⁷.

The m.e.p.c-like currents observed at ACh-loaded myocytes were induced by quantal ACh secretion from the myocyte and the subsequent activation of its surface ACh channels. External application of *d*-tubocurarine (curare) or α -bungarotoxin (α -BGT), agents known to block nicotinic ACh receptors in these myocytes^{3,4}, abolished the currents (Fig. 2a and b). The effect of curare was reversible, whereas that of α -BGT was not, which is consistent with the known actions of these agents on ACh channels. After the drug treatments, we observed a reduced current noise and a small reduction of steady inward current (4.0 ± 1.3 pA s.e.; $n = 8$), suggesting a small steady non-quantal ACh leakage, a phenomenon that occurs in mature¹⁰ and developing¹² neuromuscular junction. Treatments known to elevate the frequency of m.e.p.cs at the neuromuscular junction also increased m.e.p.c-like events in ACh-loaded myocytes. After bath application (30 min) of carbonyl cyanide *m*-chlorophenylhydrazone (CCCP, at 10 μM), an agent known to induce Ca^{2+} release from mitochondria¹³, the frequency of the event was increased to $229 \pm 25\%$ (s.e., $n = 4$) of the control (Fig. 2c). Repetitive depolarizations of the myocyte membrane also led to an immediate rise in the event frequency, presumably as a result of Ca^{2+} influx through voltage-dependent Ca^{2+} channels. After depolarizations (+80 mV for 5 ms; 300 pulses at 8 Hz), the frequency increased to $193 \pm 11\%$ (s.e., $n = 6$) of the control (Fig. 2d). The increase in cytosolic Ca^{2+} after CCCP treatment and repetitive depolarization was confirmed by monitoring Ca^{2+}

FIG. 1 Miniature endplate current-like events recorded from ACh-loaded myocytes. *a*, An example of a recording obtained from an isolated myocyte in 1-day-old *Xenopus* culture. A whole-cell, voltage-clamp recording pipette containing 10 mM of ACh was used to load ACh into the myocyte and to record myocyte membrane current ($V_c = -70$ mV, filtered at 150 Hz). Inward pulsatile currents resembling m.e.p.cs appeared within 5 min of the onset of recording. Samples of current events during the period 5–8 min and 23–25 min after the onset of recording are shown below at a higher time resolution (filtered at 2.5 kHz). Scales: 100 pA, 100 ms (left) and 400 pA, 100 ms (right). *b* and *c*, Time course of the increase in the mean frequency and mean amplitude, respectively, of m.e.p.c-like events observed from 3 separate ACh-loaded myocytes. Data points are mean values over 5-min periods. Each symbol represents data from one cell. *d*, The mean amplitude ($\square$) and mean frequency (Δ) of m.e.p.c-like events observed at 30–40 min after onset of ACh loading at different concentrations. The error bars refer to s.e.m. and $n = 10$ for each concentration.

METHODS. Procedures for preparing *Xenopus* nerve-muscle cultures have been reported²⁶⁻²⁸. These cultures were used for experiments after 1-d incubation at room temperature (20–22°C). The culture medium consisted of (v/v) 50% Leibovitz L-15 medium (Sigma), 1% fetal bovine serum (GIBCO) and 49% Ringer's solution (115 mM NaCl, 2 mM CaCl_2 , 2.5 mM KCl, 10 mM HEPES, pH 7.6). Gigaohm-seal whole-cell recording methods followed those described previously⁵⁻⁷. Recordings were made at room temperature in culture medium. The internal solution for whole-cell recording contained 150 mM KCl, 1 mM NaCl, 1 mM MgCl_2 and 10 mM HEPES buffer, pH 7.2. In all recordings, the membrane currents were monitored by a patch clamp amplifier (Axopatch 1B). Data were stored by a videotape recorder for later playback onto a storage oscilloscope (Tektronix 5113) or



an oscillographic chart recorder (Gould RS3200) and for analysis by a microcomputer. For loading ACh into myocyte, the pipette tip was first front-filled with normal internal solution for a length of about 1.5 mm, and the pipette was then back-filled with the internal solution containing ACh. This procedure reduced the activation of myocyte ACh receptors before seal formation. The characteristic diffusion time for loading small molecules into myocyte was studied by quantitative fluorometric measurements using Lucifer Yellow as a probe. Using a filling procedure similar to that already described, the characteristic loading time was estimated to be about 20 min (R. Stoop and M.-m. P., unpublished data).

levels with the Ca^{2+} -sensitive fluorescent dye, Fluo-3 (see Fig. 2 legend). The fluorescence intensity increased to $176 \pm 13\%$ (s.e., $n = 19$) and $157 \pm 14\%$ (s.e., $n = 21$) of the control after the CCCP treatment and depolarizations (Fig. 2e), respectively.

Previous studies^{14,15} have shown that exocytosis of transmitter-containing vesicles is the most likely basis for quantal transmitter secretion. Ultrastructural studies of *Xenopus* myocytes^{16,17} have demonstrated the existence of vesicles of various sizes which may serve to accumulate ACh and their exocytosis may account for the observed m.e.p.c.-like events. The main evidence indicative of ACh compartmentation came from experiments using acetylcholinesterases (AChEs). Myocytes were first loaded with 10 mM of ACh until m.e.p.c.-like events happened with a relatively high frequency. The ACh-containing pipette was then detached and a second pipette containing AChEs (300 units ml^{-1}) was used to load AChEs into the same myocyte. We observed a gradual decline in the frequency of m.e.p.c.-like events immediately after AChE loading, with no significant reduction of their mean amplitude (Fig. 3). A control second recording using a pipette containing 10 mM ACh instead of AChEs showed no change in current events over similar recording periods. This observation suggests that ACh molecules first accumulated in cellular compartments and became inaccessible to hydrolysis by AChEs. Rapid hydrolysis of the free cytosolic ACh had apparently caused immediate cessation of ACh packaging, leading to a gradual depletion of existing ACh packets, and thus to an exponential decline in the event frequency. On the other hand, if secretion were due to

direct release of cytosolic ACh through the plasma membrane, we would expect to find a reduction of the mean current amplitude, as cytosolic ACh concentration decreases through hydrolysis by AChE, without much change in the frequency. Thus ACh compartmentation seems to be the most likely mechanism for quantal secretion. This is further supported by the results of vesamicol treatment. Preincubation of the myocyte with vesamicol (at 5 μM for 30 min), an inhibitor of ACh transport into synaptic vesicles and of vesicle mobilization¹⁸⁻²⁰, resulted in a reduction of the mean amplitude of the m.e.p.c.-like events, as shown by the composite graph of the cumulative frequency of events with different amplitudes (Fig. 3d). This effect was not due to direct action of vesamicol on ACh channels, as shown by direct iontophoretic testing of ACh sensitivity (data not shown). In vesamicol-treated myocytes, the average frequency of m.e.p.c.-like events was 2.4 ± 0.6 per min (s.e., $n = 10$), which is significantly different ($P < 0.025$, Mann-Whitney U test) from that of control untreated myocytes (8.6 ± 3.2 per min, s.e., $n = 10$). Reduced vesicular uptake of ACh could cause a reduction of the current amplitude as well as an apparent reduction in frequency, as amplitudes of small events were reduced below noise level. Alternatively, the reduction of frequency may result from the effect of vesamicol on vesicle mobilization for secretion²⁰.

In the developing neuromuscular synapse, variability in the size of m.e.p.c.s probably reflects differences in the amount of ACh contained in each quantal packet⁷ which are due to an insufficient supply of ACh or incomplete ACh packaging. In

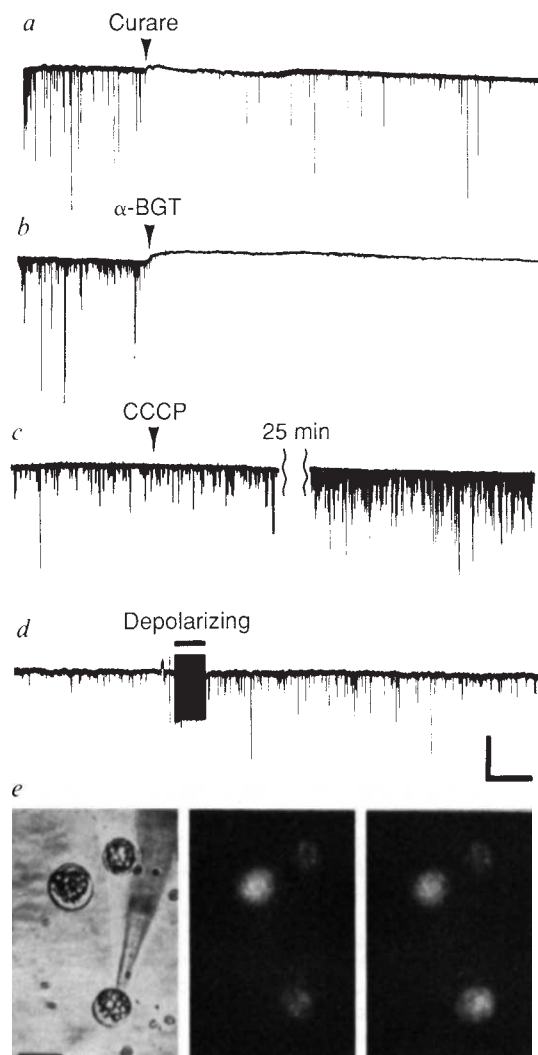
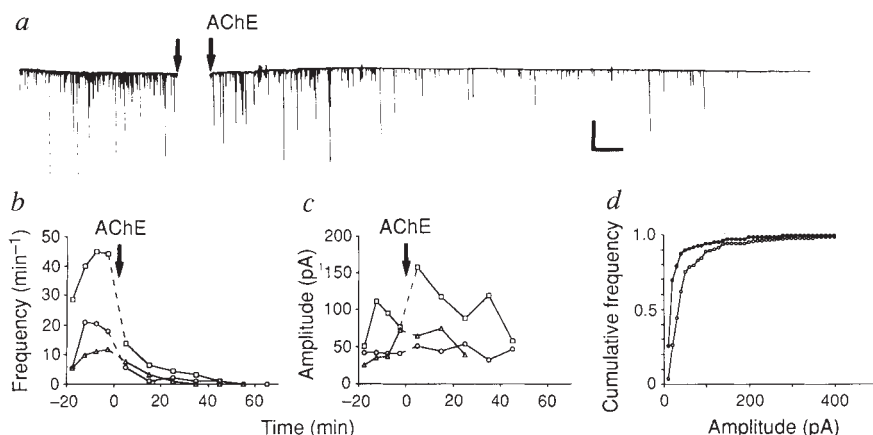


FIG. 2 Effects of pharmacological treatments and repetitive depolarizations. The ACh-loading (10 mM) and recording procedure was the same as described in Fig. 1. Treatments were performed 30 min after the onset of ACh loading. *a* and *b*, Effect of external application of pulse of culture medium containing 15 mM of *d*-tubocurarine (curare) (*a*) or 50 $\mu\text{g ml}^{-1}$ of α -bungarotoxin (α -BGT) (*b*), respectively, by pressure-ejection through a micropipette positioned near the myocyte. *c*, Effect of treatment with CCCP (Sigma; 10 μM). For display the trace over a 25 min period was omitted. Scale: 100 pA, 2.5 min. *d*, Effect of applying 300 depolarizing voltage steps (+80 mV, 5 ms, 8 Hz) to the myocyte. The external Ca^{2+} concentration was adjusted to 10 mM. *e*, Effect of repetitive depolarization (same as *d*) on the myocyte Ca^{2+} level was monitored directly with Ca^{2+} -sensitive-dye Fluo-3. Bright-field photograph of the culture (left) shows the myocyte recorded with a whole-cell pipette containing 10 mM of ACh. Bar, 30 μm . The culture medium contained 10 mM Ca^{2+} . Fluorescence photographs showed images of the myocytes before (centre) and after (right) application of 300 pulses of depolarization steps (same as *d*). Note the increase in fluorescence in the recorded myocyte.

METHODS. The Ca^{2+} -sensitive dye Fluo-3-pentaacetyloxymethyl ester²⁹ (Fluo-3-AM; Molecular Probes) was used to monitor changes in cytosolic Ca^{2+} concentration in the myocyte. Myocytes were incubated with Fluo-3-AM (2 μM) for 30 min at room temperature and extensively washed with fresh culture medium before examination with an inverted fluorescence microscope (Diaphot). A water-immersion objective ($\times 40$) was used with an intensified CCD camera (Quantex, QX-100) operating at a fixed gain. Excitation and emission fluorescence spectra used typical fluorescein filters (excitation wavelength 420–460 nm and barrier filter 520–560 nm) coupled with a neutral density filter to prevent phototoxicity. Fluorescent images were analysed by means of a digital image processor (Imaging Technology 151). The images were 'grabbed' as an array of 512×512 pixels with 256 grey levels. Typically 32 or 64 frames were averaged for quantitative analysis. The fluorescence intensity was obtained by sampling an area covering the entire myocyte. The background fluorescence, measured at cell-free areas, was subtracted from each measurement. The extent of photobleaching of Fluo-3 fluorescence was less than 5% for a typical experiment. The non-linearity of the CCD camera is negligible under the resolution of the present study. Changes in the fluorescence intensity of Fluo-3 were measured by obtaining the ratio of intensity after the CCCP or depolarization treatment to that before the treatment.

FIG. 3 *a*, Effect of acetylcholinesterases (AChE) on m.e.p.c.-like events in ACh-loaded myocytes. The ACh-loading and recording procedure was the same as described in Fig. 1. The ACh-containing recording pipette was detached from the myocyte 20 min after the onset of recording (at the time marked by the first arrow), and a new recording pipette containing 300 units ml^{-1} AChEs (Sigma) was used to load AChE into the myocyte (at time marked AChE) and to record the membrane current. Scale, 300 pA and 2.5 min. *b* and *c*, The mean frequency and mean amplitude of m.e.p.c.-like events, respectively, at various times before and after loading AChEs. Data are from 3 separate cells, each represented by the same symbol. Each point represents mean values determined over 5- or 10-min periods. *d*, The effect of pretreatment with vesamicol (30 min) on the amplitude distribution of m.e.p.c.-like events. Events during the period 30–35 min after the onset of ACh-loading (10 mM) were analysed. The cumulative frequency refers to percentage of total events below different



amplitudes for vesamicol-treated (●) and control (○) myocytes. Total number of events sampled were 95 and 348 (10 cells each), respectively.

this study, highly variable amplitudes of m.e.p.c.-like events in ACh-loaded myocytes were found for as long as the recording was made (up to 2 h). But after 30 min of ACh loading, events with amplitudes of up to 1 nA were observed. These events correspond to simultaneous opening of ~ 500 ACh channels. Considering that two ACh molecules are required to open an ACh channel and that substantial loss of secreted ACh molecules is likely, then the amount of ACh contained within the packets responsible for these large events could be a few thousand, reaching the range of 5,000 to 10,000 estimated for ACh quanta at mature neuromuscular functions²¹. As the mean amplitude of m.e.p.c.-like events in myocytes was of the order of 100 pA, the number of ACh molecules in an average quantal packet is likely to be of the order of 1,000. Examination of electron microscopic images^{16,17} of these myocytes showed a population of vesicles with an average diameter of 50 nm ($n=40$, range

24–80 nm). If passive diffusional filling of these vesicles had occurred, the number of ACh molecules in the vesicle will be about 500 (for 10 mM cytosolic ACh). On the other hand, it is unlikely that charged ACh molecules can diffuse across vesicular membrane effectively without assistance, as suggested by the low ACh leakage-induced membrane current (see above). Furthermore, the constancy of average amplitudes following the AChE loading suggests no substantial leakage of the accumulated ACh from the compartments back to the cytosol. Whether or not active accumulation of ACh occurs with the help of a specific ACh transporter remains to be investigated.

In addition to spontaneous secretion, we have also examined evoked ACh release from ACh-loaded myocytes by applying a brief depolarizing step to the voltage-clamped myocyte. Inward currents of various amplitudes following the depolarizations were observed, together with many failures in the evoked

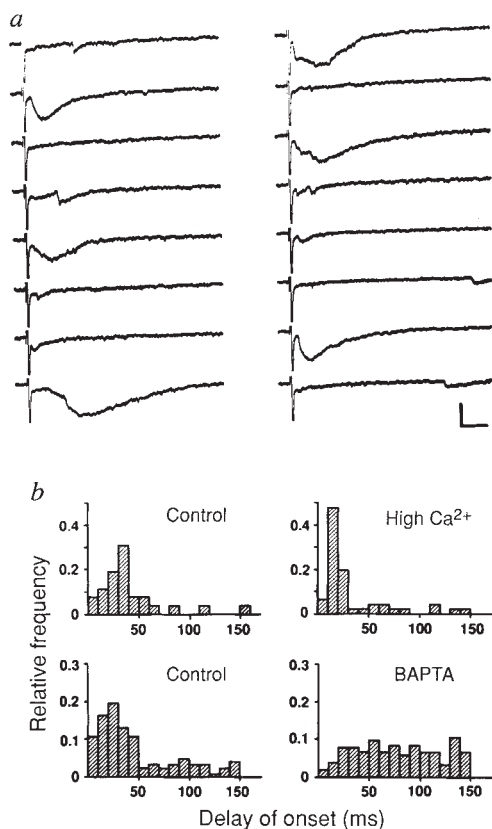


FIG. 4 Depolarization-evoked ACh release from ACh-loaded myocytes. The ACh-loading and recording procedure was the same as described in Fig. 1 legend. *a*, Sixteen consecutive traces of myocyte membrane currents following application of depolarizing steps (+80 mV, 5 ms, 1 Hz) to a voltage-clamped myocyte ($V_c = -70$ mV), 30 min after the onset of ACh-loading. External Ca^{2+} concentration was 10 mM. Scales, 100 pA and 100 ms. *b*, Upper graphs: Normalized distribution of onset times of inward current events recorded from one myocyte following each of 300 depolarizing voltage steps (same as in *a*), in medium containing normal (2 mM, control) or high (10 mM) concentration of Ca^{2+} , respectively. The onset of the voltage step was set as time 0. Relative frequency refers to the % of the total events observed with different onset times. Lower graphs: Distribution of onset times for control and BAPTA-loaded myocytes (10 cells each). Normalized distribution of onset times was determined for each cell and the graphs represent the average frequency distribution. Note the random distribution of current events in BAPTA-loaded cells.

METHODS. BAPTA (1,2-bis(2-aminophenoxy)ethane- N,N,N',N' -tetraacetic acid³⁰, Sigma; 5 mM) was loaded into the myocyte by means of a whole-cell recording pipette, which was filled with normal internal medium supplemented with 2 mM Ca^{2+} and 10 mM ACh. The buffered Ca^{2+} concentration in the myocyte was estimated to be 10^{-7} M.

response (Fig. 4a). Elevation of external Ca^{2+} from the normal level of 2 mM to 10 mM significantly reduced the rate of failure in evoking ACh release. The onset times of the current events following depolarizing steps were also significantly reduced in high Ca^{2+} medium (Fig. 4b, upper graph). When the myocyte cytosolic Ca^{2+} was buffered at a low level (10^{-7} M) by loading a Ca^{2+} buffer, BAPTA, the depolarization-evoked secretion events were abolished (Fig. 4b, lower graphs), leaving random spontaneous events occurring at a reduced frequency. These results are consistent with a Ca^{2+} -regulated evoked secretion, triggered by the Ca^{2+} influx through voltage-dependent calcium channels in the myocyte membrane. It is also clear that the excitation-secretion coupling in the myocyte was significantly weaker than that of the neuromuscular junction, where the delay of onset of evoked responses is less than

1 ms and the failure occurs only at much lower level of external Ca^{2+} .

In summary, we have shown that myocytes loaded with exogenous ACh exhibited spontaneous quantal ACh secretion and depolarization-evoked ACh release, with a Ca^{2+} -dependence reminiscent of ACh secretion from presynaptic nerve terminals. The ACh appeared to accumulate in cytoplasmic compartments before secretion. Thus, with sufficient ACh supply in the cytosol, Ca^{2+} -regulated exocytotic secretion of ACh can occur in a non-neuronal cell. In comparison with neuronal secretion, however, excitation-secretion coupling in the myocyte is weak. Future work of introducing synapse-specific components²²⁻²⁵ into ACh-loaded myocytes should reveal their individual contribution to the assembly and the regulation of an efficient mechanism for transmitter secretion. □

Received 11 August; accepted 4 September 1992.

- Katz, B. *The Release of Neural Transmitter Substances* (Thomas, Springfield, 1969).
- Kandel, E. R., Schwartz, J. H. & Jessel, T. *Principles of Neural Science* 3rd edn (Elsevier, Amsterdam, 1991).
- Orida, N. & Poo, M.-m. *Nature* **275**, 31-35 (1978).
- Poo, M.-m. *Nature* **295**, 332-334 (1982).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 80-100 (1981).
- Young, S. H. & Poo, M.-m. *Nature* **304**, 161-163 (1983).
- Evers, J., Laser, M., Sun, Y., Xie, Z. & Poo, M.-m. *J. Neurosci.* **9**, 1523-1539 (1989).
- Morel, N., Israel, M. & Manaranche, R. *J. Neurochem.* **30**, 1553-1557 (1978).
- Israel, M., Dunant, Y. & Manaranche, R. *Progr. Neurobiol.* **13**, 237-275 (1979).
- Katz, B. & Miledi, R. *Proc. R. Soc. Lond. (Biol.)* **196**, 59-72 (1977).
- Poulain, B., Baux, G. & Tauc, L. *J. Physiol., Paris* **81**, 270-277 (1986).
- Sun, Y. & Poo, M.-m. *J. Neurosci.* **5**, 634-642 (1985).
- Tabti, N., Lupa, M. T., Yu, S. P. & Thesleff, S. *Acta physiol. Scand.* **128**, 429-426 (1986).
- Whittaker, V. B., Essman, W. B. & Dowe, G. H. C. *Biochem. J.* **128**, 833-846 (1972).
- Heuser, J. E. *et al. J. Cell Biol.* **81**, 275-300 (1979).
- Takahashi, T., Nakajima, Y., Hirose, K., Nakajima, S. & Onodera, K. *J. Neurosci.* **7**, 473-481 (1987).

- Buchanan, J., Sun, Y. & Poo, M.-m. *J. Neurosci.* **9**, 1544-1554 (1989).
- Marshall, I. G. & Parsons, S. M. *Trends Neurosci.* **10**, 1174-1177 (1987).
- Marshall, I. G., Prior, C. & Searl, T. *J. Physiol., Lond.* **424**, 52P (1990).
- Enomoto, K. *Eur. J. Pharmacol.* **147**, 209-215 (1988).
- Kuffler, S. W. & Yoshikami, D. *J. Physiol., Lond.* **251**, 465-482 (1975).
- De Camilli, P., Benfenati, F., Valtora, F. & Greengard, P. *A. Rev. Cell Biol.* **6**, 433-460 (1990).
- De Camilli, P. & Jahn, R. *A. Rev. Physiol.* **52**, 625-645 (1990).
- Südhof, T. C. & Jahn, R. *Neuron* **6**, 665-677 (1991).
- Trimble, A. W., Linial, M. & Scheller, R. *Rev. Neurosci.* **14**, 93-122 (1991).
- Spitzer, C. C. & Lamborghini, H. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1641-1645 (1976).
- Anderson, M. J., Cohen, M. W. & Zorychta, E. *J. Physiol., Lond.* **268**, 731-758 (1977).
- Tabti, N. & Poo, M.-m. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 137-153 (MIT, Cambridge, 1991).
- Minta, A., Kao, J. P. Y. & Tsien, R. Y. *J. Biol. Chem.* **264**, 8171-8178 (1989).
- Adler, E. M., Augustine, G. J., Duffy, S. N. & Charlton, M. P. *J. Neurosci.* **11**, 1496-1507 (1991).

ACKNOWLEDGEMENTS We thank R. Girod for discussion. This work was supported by grants from the NIH, the NSF and the ONR. Y.D. was supported by a Howard Hughes Predoctoral Fellowship.

Patterns of elevated free calcium and calmodulin activation in living cells

Klaus Hahn*, Robbin DeBiasio & D. Lansing Taylor

Center for Light Microscope Imaging and Biotechnology, and Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, USA

THE temporal and spatial dynamics of intracellular signals and protein effectors are being defined as a result of imaging using fluorescent reagents within living cells¹⁻⁵. We have described a new class of fluorescent analogues² termed optical biosensors⁶, which sense chemical or molecular events through their effects on protein transducers⁷. One example of this new class of indicators is MeroCaM, an environmentally sensitive fluorophore which when it is attached to calmodulin reflects the activation of calmodulin by calcium *in vitro*². We report here that the rise in free calcium and MeroCaM activation occur in the same period during serum stimulation of quiescent fibroblasts. MeroCaM activation also correlates with the spatial pattern of increased free calcium and the contraction of transverse fibres during wound healing^{1,8-10}. Finally, migrating fibroblasts in the later stages of wound-healing exhibit an increasing gradient of free calcium and MeroCaM activation from the front to the rear.

We investigated two cellular events to test prevailing hypotheses, derived from solution biochemistry^{11,12}, concerning calmodulin's role in calcium-linked activation of cell functions. First, we correlated calcium transients and MeroCaM activation during serum stimulation of quiescent cells to test whether

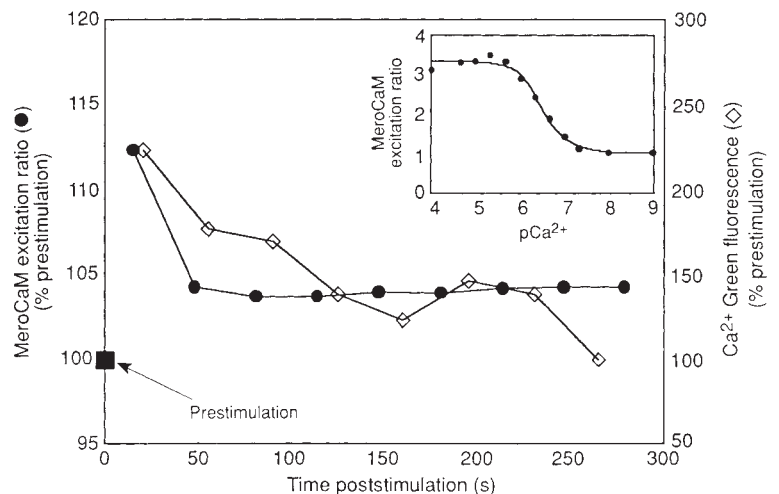
calmodulin activation is coupled to calcium elevation. Second, we mapped the spatial variation of free calcium concentration and MeroCaM activation in polarized, motile cells during wound-healing to explore the potential role of calcium-calmodulin signalling in regulating myosin II-based contraction of transverse fibres^{1,8-10}.

MeroCaM undergoes a calcium-dependent, threefold change in excitation ratio *in vitro* (Fig. 1, inset) at long wavelengths (608/532 nm) that minimize cell damage and interference from cellular autofluorescence² (Fig. 1). The biosensor's suitability for ratio imaging is valuable for quantitative interpretation of fluorescence in living cells¹³. Ratio imaging eliminates effects of variable intracellular biosensor concentration, cellular path-length and other factors¹³. Since MeroCaM was characterized *in vitro*², an optimized preparative procedure has produced MeroCaM containing 80-90% of a single fluorescent product and less than 5% unlabelled calmodulin. Titration of this preparation with myosin light-chain kinase produced a slight change in excitation ratio (0.15-fold), indicative of binding to this target protein with $K_d < 10$ nM, and demonstrating that the fluorescence response of MeroCaM is not greatly affected by binding to a target protein.

Investigation of MeroCaM behaviour in live cells directly tests the hypothesis that calmodulin is activated by increased cytoplasmic free calcium. MeroCaM response was studied during serum stimulation of quiescent Swiss 3T3 fibroblasts. Such stimulation results in a rapid increase of intracellular free calcium followed by a return to almost initial values within several minutes¹⁴⁻¹⁶. Figure 1 shows experiments in which quiescent fibroblasts were injected with MeroCaM or the fluorescent calcium indicator, Calcium Green. The results verified the previously reported calcium response to stimulation¹⁴⁻¹⁷ and the large cell-to-cell heterogeneity of response^{5,14,15,18}, and demonstrated the responsiveness of MeroCaM to calcium *in vivo*. Serum stimulation produced a transient increase in the average MeroCaM excitation ratio, indicating calmodulin activation in these cells.

* Present address: Department of Neuropharmacology, Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, USA.

FIG. 1 Responses of MeroCaM and the fluorescent calcium indicator Calcium Green during serum stimulation of quiescent Swiss 3T3 fibroblasts. Calcium Green was microinjected into quiescent Swiss 3T3 fibroblasts and the average cellular fluorescence determined periodically for each cell during stimulation with serum (excitation, 470 nm; emission, 530 nm). A representative trace from one cell is shown. In all 19 cells tested, serum induced an initial increase of 25–180% in Calcium Green fluorescence. Fluorescence then dropped towards prestimulation values, with additional smaller fluorescence maxima observed in some cases. These results demonstrated the time course of the calcium transients as a reference for MeroCaM studies. Cells were also microinjected with MeroCaM and the average fluorescence excitation ratio was determined periodically for each cell during stimulation (605/530 nm; emission, 635 nm). In 4 out of 10 cells examined, the average cellular fluorescence ratio increased 5–20% at the first time point after stimulation, indicative of activation of MeroCaM. Four cells showed an increase of less than 5%, and in one the fluorescence only decreased. After the initial poststimulation time point, the ratio always decreased sharply, then reached a plateau above prestimulation levels or continued to drop. The fluorescence ratio of MeroCaM varied by at most 5% over 100 seconds following stimulation with low-serum control medium. We cannot explain the drop in fluorescence of a subpopulation of cells. Controls rule out simple photobleaching artefacts. Cells were prepared



and stimulated according to published methods¹⁸. All images used were acquired, analysed and displayed with the Multimode microscope (Biological Detection Systems, PA). A cooled CCD camera was employed with 2–3 second exposure. Inset, the calcium-dependence of MeroCaM's excitation ratio, from ref. 2.

Imaging of individual cells revealed spatial aspects of calmodulin activation not accessible using other approaches. Figure 2 demonstrates the temporal and spatial heterogeneity of calmodulin activation in stimulated cells. The ratio-imaging technique normalizes local probe concentration^{5,13}, so that the signals shown represent MeroCaM activation. In a number of cells, calmodulin activation occurred at the cell periphery immediately after stimulation, and then propagated towards the nucleus. The level of calmodulin activity in these areas fell and rose during movement to the perinuclear region. These spatially heterogeneous responses could result from subcellular variations in calcium concentration, or from effects of target proteins on calmodulin's calcium affinity¹². It has previously been demonstrated that serum stimulation of quiescent fibroblasts causes stress fibres to contract and condense in the perinuclear region¹⁸. Calmodulin target enzymes such as myosin light-chain kinase are thought to regulate myosin II (refs 19–21) and so may condense with the stress fibres.

We also analysed MeroCaM in highly polarized cells undergoing wound-healing to test whether myosin-II based contractions^{1,8–10} *in vivo* involve calmodulin activation. A wound was introduced in a monolayer of Swiss 3T3 fibroblasts and cells along the wound's edge were microinjected with MeroCaM alone or in combination with the calcium indicator (legends to Figs 3 and 4). Figure 3 shows the temporal and spatial dynamics of calmodulin activation monitored early in the wound-healing process. Calmodulin activation was greatest in the lamellae where contraction of transverse fibres made the anterior of the cell narrower and induced detachment from neighbouring cells^{1,8–10}. Figure 4a and b shows that elevated free calcium correlates with MeroCaM activation in the lamellum, where transverse fibres contract while transporting to the perinuclear region¹. The localized rise in free calcium and activation of calmodulin in this region indicated that contraction of the transverse fibres is regulated spatially by calcium and calmodulin¹⁹. Figure 4c and d shows the existence of an increasing gradient

FIG. 2 Spatial distribution of calmodulin activation in fibroblasts during serum stimulation. Quiescent Swiss 3T3 fibroblasts were microinjected with MeroCaM according to ref. 25 and the biosensor's excitation ratio was imaged during serum stimulation. The spatial distribution of calmodulin activation is shown. Higher excitation ratios, and hence greater calmodulin activation, correspond to warmer colours (red maximal). The difference between the highest and lowest values corresponds to a 2.5-fold difference in excitation ratio. Free calcium concentration of quiescent fibroblasts is below 0.1 μ M and stimulation with serum induces heterogeneous responses of up to several micromolar at peak values^{14–16}. This representative sequence of images illustrates movement of activated regions towards the nucleus over time, most clearly visible in the cell on the upper left. Estimated intracellular concentrations were Calcium Green, 10 μ M; MeroCaM, 10–30 μ M.

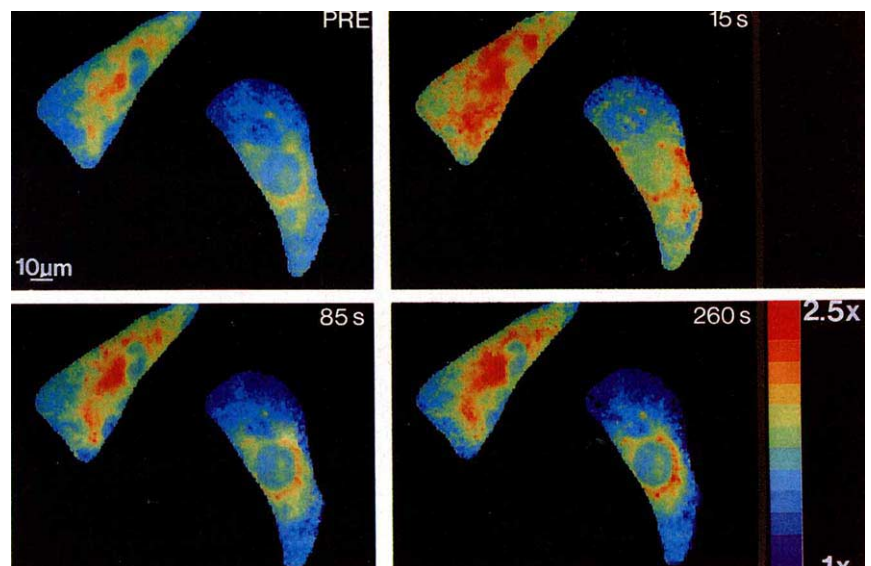
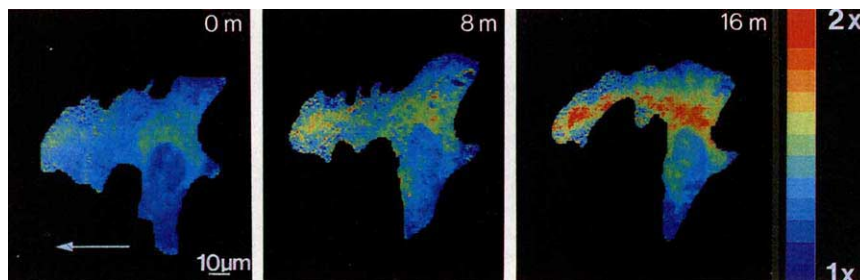


FIG. 3 Subcellular localization of calmodulin activation in fibroblasts during the lamellar contraction phase of wound-healing. A 'wound' was introduced in a monolayer of Swiss 3T3 fibroblasts⁸⁻¹⁰ and cells at the edge of the wound were microinjected with MeroCaM. The excitation ratio of MeroCaM was imaged in time-lapse in individual cells as they contracted during the separation from neighbouring cells. In all actively contracting cells (7 of 14 cells examined), calmodulin activity was greatest in the leading lamellae, regions undergoing strong contraction of transverse fibres that contain actin and myosin II⁸⁻¹⁰. Note the increased activation of MeroCaM in the time series as the leading lamellum is constricted by the contraction. As in Fig. 2, warmer colours correspond to greater MeroCaM ratios, and hence to greater calmodulin activation. There was a 2.5-fold difference between regions of lowest activity (blue) and highest activity (red) (see Fig. 1). The arrow indicates the direction of orientation and movement. Cells were microinjected 5 h after wounding and images were taken 30 min later



at 8-min intervals. A control experiment to determine the distribution of MeroCaM was done by co-injecting MeroCaM with the volume indicator fluorescein-dextran (M_r 10K). The ratio image (MeroCaM/fluorescein-dextran) using the isosbestic point of MeroCaM demonstrated a relatively uniform distribution of MeroCaM, except where there was extensive contraction which is consistent with earlier studies with fluorescent analogues of calmodulin^{25,26}.

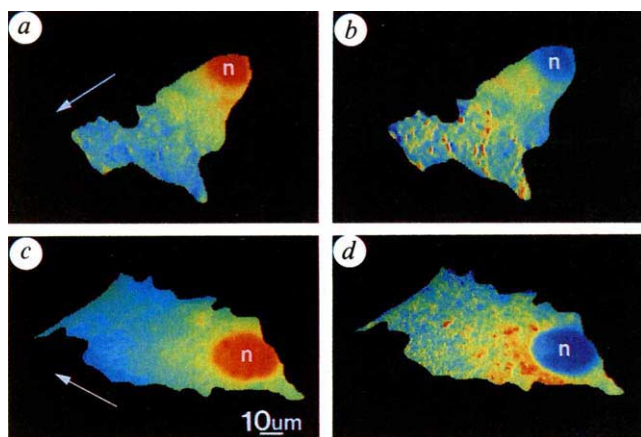


FIG. 4 Subcellular localization of increases in free calcium ions and MeroCaM activation at single time points in fibroblasts during two stages of wound-healing^{1,8-10}. Cells were co-injected with a mixture of Calcium Green/Dextran 10K, the calcium-insensitive volume indicator Cascade Blue/Dextran 10K (both from Molecular Probes), and MeroCaM to produce ratio images mapping regions of elevated free calcium (Calcium Green/Cascade Blue) and MeroCaM activation (excitation 605/530 nm, emission 635 nm). *a* and *b*, One time point during the contraction of transverse fibres demonstrates the similar pattern of elevated free calcium and MeroCaM activation in the lamellum during the lamellar contractile phase of wound-healing. The pattern of elevated free calcium (*a*) and MeroCaM activation (*b*) were very similar in 12 cells investigated. *c* and *d*, Increasing gradient of free calcium (*c*) and MeroCaM activation (*d*) from the front of the cell to the rear of the lamellum of a migrating cell. This gradient was detected in all 12 polarized cells. Warmer colours represent both elevated free calcium and MeroCaM activation. Arrows indicate the direction of orientation and movement; *n*, nucleus. The relatively high and sustained free calcium concentration in the nucleus is detected in many activated cells and is not understood, but may reflect nuclear activation¹⁵. MeroCaM activation is not always correlated with elevated free calcium in the nucleus. A quantitative ratio method is currently being developed to take advantage of the relatively high fluorescence signal from Calcium Green even at low free-calcium-ion concentrations. Neutral dextrans will be used as optimal carriers of ion indicators to maintain the indicator in the free cytoplasmic compartment^{5,13}.

of free calcium and calmodulin activation from the front to the rear of migrating cells during later stages of wound-healing. The elevated free calcium in the rear of locomoting fibroblasts is comparable to that originally demonstrated in amoebae²² and then in eosinophils²³. The gradient of free calcium may be involved in regulating the polarity of actin and myosin II assembly at the front of the migrating cell^{8,10}, as well as in the rearward transport and contraction of the transverse fibres. The myosin II-based transverse fibres contract maximally and then

disappear as they transport into the perinuclear region^{1,3,8-10}. The mottled appearance of the MeroCaM activation (Fig. 4*d*), especially in the perinuclear region, may indicate a concentration of calmodulin during the maximal contraction of transverse fibres.

The ability to study the temporal-spatial dynamics of chemical and molecular processes in individual cells with a variety of reagents and modes of light microscopy will be important in defining the mechanisms of cell functions such as motility²⁴. □

Received 8 June; accepted 4 September 1992.

1. Kolega, J. & Taylor, D. L. *Curr. Top. Membr.* **38**, 187-206 (1991).
2. Hahn, K. M., Waggoner, A. S. & Taylor, D. L. *J. biol. Chem.* **265**, 20335-20345 (1990).
3. Taylor, D. L. *et al. Cell Structure and Function by Spectrofluorimetry* (eds Kohen, E. & Hirschberg, J. G.) 297-313 (Academic, New York, 1989).
4. Tsien, R. Y. *Meth. Cell Biol.* **30**, 127-153 (1989).
5. Bright, G. R., Whitaker, J. E., Haugland, R. P. & Taylor, D. L. *J. cell. Physiol.* **141**, 410-419 (1989).
6. Hahn, K. M. *et al. Fluorescent Probes for Biological Function in Living Cells* (eds Mason, W. T. & Relf, G.) (Academic, London, in the press).
7. Adams, S. R., Harootyan, A. T., Buechler, Y. J., Taylor, S. S. & Tsien, R. Y. *Nature* **349**, 694-697 (1991).
8. Fisher, G. W., Conrad, P. A., DeBiasio, R. L. & Taylor, D. L. *Cell Motil. Cytoskel.* **11**, 235-247 (1988).
9. Conrad, P. A., Nederlof, M. A., Herman, I. M. & Taylor, D. L. *Cell Motil. Cytoskel.* **14**, 527-543 (1989).
10. DeBiasio, R. L., Wang, L., Fisher, G. W. & Taylor, D. L. *J. Cell Biol.* **107**, 2631-2645 (1988).
11. Mills, J. S., Walsh, M. P., Nemcek, K. & Johnson, J. D. *Biochemistry* **27**, 991-996 (1988).
12. Cohen, P. & Klee, C. (eds), *Calmodulin* (Elsevier, New York, 1988).
13. Bright, G. R., Fisher, G. W., Rogowska, J. & Taylor, D. L. *Meth. Cell Biol.* **30**, 157-190 (1989).

14. Byron, K. L. & Villereal, M. L. *J. biol. Chem.* **264**, 18234-18239 (1989).
15. Tucker, R. W. & Fay, F. S. *Eur. J. Cell Biol.* **51**, 120-127 (1990).
16. McNeill, P. L., McKenna, M. P. & Taylor, D. L. *J. Cell Biol.* **101**, 372-379 (1985).
17. Berridge, M. J. *J. biol. Chem.* **265**, 9583-9586 (1990).
18. Giuliano, K. *et al. Cell Motil. Cytoskel.* **16**, 14-21 (1990).
19. Sellers, J. R. & Adelstein, R. S. *The Enzymes* **XVIII**, 381-418 (1987).
20. Sellers, J. R. *Curr. Opin. Cell Biol.* **3**, 98-104 (1991).
21. Kamm, K. E. & Stull, J. T. *A. Rev. Pharmac. Toxicol.* **25**, 593-620 (1985).
22. Taylor, D. L., Blinks, J. R. & Reynolds, G. T. *J. Cell Biol.* **86**, 599-607 (1980).
23. Brundage, R. A., Fogarty, K. E., Tuft, R. A. & Fay, F. S. *Science* **254**, 703-706 (1991).
24. Taylor, D. L., Nederlof, M., Lanni, F. & Waggoner, A. S. *Am. Sci.* **80**, 322-335 (1992).
25. Luby-Phelps, K., Lanni, F. & Taylor, D. L. *J. Cell Biol.* **101**, 1245-1256 (1985).
26. Zavortink, M. M., Welsh, M. J. & McIntosh, J. R. *Expl Cell Res.* **149**, 375-385 (1983).

ACKNOWLEDGEMENTS. We thank A. Waggoner, J. Montibeller, D. Patek, K. Giuliano, G. Fisher, J. Myers, P. Southwick, A. Koretsky, B. Wise and C. Klee for help and advice. This work was supported by the NIH and NSF.

Inactivation of the sarcoplasmic reticulum calcium channel by protein kinase

Jixin Wang & Philip M. Best*

Department of Physiology and Biophysics and the School of Medicine, University of Illinois, 524 Burrill Hall, 407 S. Goodwin Avenue, Urbana, Illinois 61801, USA

* To whom correspondence should be addressed

THE ryanodine receptor protein of skeletal muscle sarcoplasmic reticulum (SR) membranes is a calcium ion channel which allows movement of calcium from the SR lumen into the cytoplasm during muscle activation¹⁻⁵. Gating of this channel is modulated by a number of physiologically important substances including calcium. Interestingly, calcium has both activating and inactivating effects which are concentration- and tissue-specific. In skeletal muscle, calcium-dependent inactivation of calcium release occurs at concentrations reached physiologically, suggesting that calcium may modulate the release process by a negative feedback mechanism⁶⁻⁹. To determine the cellular mechanism responsible for calcium-dependent inactivation, we have investigated the ability of protein phosphorylation to affect single channel gating behaviour using the patch clamp technique. Here we demonstrate that the ryanodine receptor protein/calcium release channel of skeletal muscle SR is inactivated under conditions permissive for protein phosphoryla-

TABLE 1 Effect of ATP and kinase inhibitors on channel gating

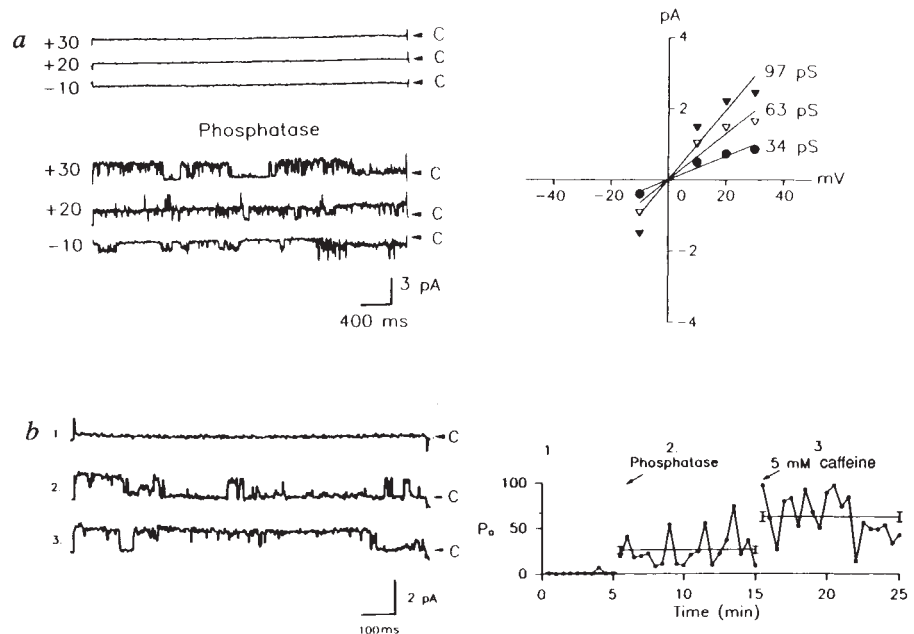
Treatment	Number of channels inhibited	Number of channels unaffected
ATP	8	1
ATP + CaM kinase II inhibitor (273-302)	1	7
ATP + control peptide (284-302)	6	1
ATP + PKC inhibitor	4	0

tion. This inactivation is reversed by the application of phosphatase and prevented by a peptide inhibitor specific for calcium/calmodulin-dependent protein kinase II. The results provide evidence for an endogenous protein kinase which is closely associated with the ryanodine receptor protein and regulates channel gating.

Single channel currents were recorded from excised patches of SR membrane, with the cytoplasmic face oriented toward the lumen of the pipette^{10,11}. When the pipette solution contained 0.12 μ M calcium and 1 mM ATP to maximize calcium-dependent inactivation of SR calcium channels^{8,9}, most patches showed no opening events, suggesting either that the patches contained no channels or that channel-open probability (P_o) was very low (Fig. 1). These silent patches were exposed by pipette perfusion to a solution containing protein phosphatase and 0 mM ATP. Channel activity appeared within a few minutes of the start of perfusion (Fig. 1a). In a total of 12 patches exposed to phosphatase, the average P_o increased from 0.02 ± 0.01 to 0.70 ± 0.07 ($\pm$ s.e.m., $n = 12$, $P < 0.01$). In each case,

FIG. 1 Activity of SR release channel after activation by phosphatase. **a**, Excised patch recording before (top) and after (bottom) application of 25 units ml^{-1} of nonspecific acid phosphatase (Calbiochem) to the patch by pipette perfusion (C, closed state). The pipette solution contained 1 mM Mg-ATP and 0.12 μ M Ca^{2+} . The phosphatase stock was dialysed against pipette solution before use. **I-V** relationships for three conductance levels seen after application of phosphatase are shown on the right. The largest conductance state was not seen in every patch and, when present, was rare ($< 5\%$ of total recording time). **b**, Time course of change in P_o from a single patch during exposure to phosphatase and then to 5 mM caffeine (intervals 1-3). Membrane initially exposed to 1 mM ATP and 0.12 μ M Ca^{2+} to promote inactivation before patch was made. Each point represents the P_o for a 30-s period calculated by averaging the P_o from six 1-s-long data records taken every fifth second during that period. Arrows indicate time of pipette perfusion with a new solution. Horizontal lines show the grand average P_o and vertical bars the s.e. for each interval. Representative data traces from indicated time periods are shown to the left. Currents recorded during voltage clamp steps from 0 to 20 mV.

METHODS. Extruded blebs of frog skeletal muscle SR membrane were prepared and patched as described^{10,11}. Briefly, the surface membrane of single fibres from frog semitendinosus muscle was removed manually, and the 'skinned' fibre stimulated to contract against zero load by exposure to a calcium-containing solution. As the fibre contracted, blebs of membrane were extruded from the fibre interior. Excised patches were made from this extruded membrane. High-resistance pipettes (20-30 M Ω) were used to increase the frequency of obtaining single channel patches. Bath and pipette solutions always contained (in mM) 111 sodium glutamate, 2 EGTA, 1.39 calcium glutamate, 1 MgSO_4 , and 25 MOPS buffer pH 7.2, at room tem-



perature. Free calcium was calculated to be 0.12 μ M. When ATP was present, additional Mg^{2+} was added to the pipette solution to maintain both Mg^{2+} and Mg-ATP at 1 mM. Under these conditions, currents through the SR Cl^- and K^+ channels are minimized and the major charge carrier through the SR Ca^{2+} channel is Na^+ at both positive and negative voltages. The membrane was held at 0 mV and stepped to the voltage levels indicated. Single channel recording and analysis was done with commercially available software that used a half-maximum threshold detection criterion (Pclamp, Axon Instruments). Data were filtered at 500 Hz and sampled at 1 kHz. Pipette perfusion was accomplished by inserting two pieces of quartz tubing down the shank of the pipette into the tip. New solution was drawn into the pipette through one tube (pulled to a narrow diameter and placed near the pipette tip) while suction was applied to the second tube (of larger diameter placed farther up the shank).

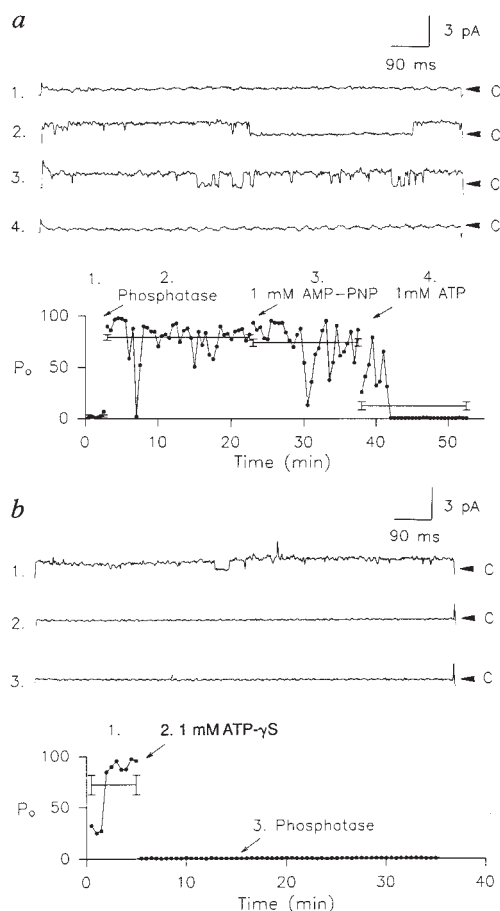


FIG. 2 Time course of changes in P_o from a single patch after pipette perfusion with ATP and its analogues. *a*, Effect of ATP and AMP-PNP on channel activity. The membrane was exposed to a solution containing 1 mM ATP and $0.12 \mu\text{M Ca}^{2+}$ to promote inactivation before the patch was formed. *b*, Effect of ATP- γ S. The SR membrane was treated with phosphatase before the patch was made to activate channels.

phosphatase-activated channels displayed multiple open states (not all of which were seen in every patch), as previously described for the calcium-release channel^{5,10}. These results suggest that channel gating had been inhibited in the silent patches and that removal of phosphate from some protein in the patch led to an increase in channel P_o ¹². Removal of ATP without adding phosphatase had no effect on channel activity under these conditions. The effect of phosphatase was also seen in the presence of ATP, although the results were more variable and not every patch showed channel activation.

The channels activated by phosphatase treatment were identified as being SR calcium release channels because they were activated by caffeine, an agent that specifically affects SR calcium channel gating^{10,13}. After obtaining a patch, a quiescent channel was activated by perfusion of the pipette with phosphatase (Fig. 1*b*). The same patch was then exposed to caffeine by a second exchange of the pipette solution. The phosphatase-activated channel showed an additional increase in P_o after caffeine exposure. In three experiments, the average P_o increased from 0.42 ± 0.08 after phosphatase activation to 0.74 ± 0.06 after caffeine exposure ($P < 0.02$). Neither the conductance nor voltage dependency of the channels was changed by caffeine (data not shown).

Because phosphatase activated previously silent channels, we next tested the ability of ATP and several ATP analogues to diminish the open probability (P_o) of active channels (Fig. 2). Channel activity was monitored during an initial control period

(1 mM ATP, $0.12 \mu\text{M Ca}^{2+}$) and then after sequential perfusion of the pipette with phosphatase (0 mM ATP), the non-hydrolysable ATP analogue AMP-PNP (1 mM) and ATP (1 mM). During the control period the patch showed little channel activity ($P_o = 0.02$). Perfusion with phosphatase led to a sustained increase in channel activity ($P_o = 0.79$), consistent with the data shown in Fig. 1. Subsequent perfusion of the pipette with a phosphatase-free solution containing 1 mM AMP-PNP did not affect channel gating significantly. But when the patch was re-exposed to 1 mM ATP (no phosphatase, $0.12 \mu\text{M Ca}^{2+}$), channel activity was reduced to a level ($P_o = 0.12$) approaching that before phosphatase exposure. Thus the inhibitory effect of ATP on channel gating requires ATP hydrolysis. Although this was the only patch in which we were successful in making these three consecutive changes in the pipette solution, the effects of AMP-PNP and ATP were confirmed individually in 13 other patches. In separate experiments, the pipette was perfused with 1 mM ATP- γ S following activation of the patch by phosphatase (Fig. 2*b*). Exposure to ATP- γ S led to an inhibition of channel activity which was not reversed by subsequent application of phosphatase ($n = 5$). This is consistent with the known inability of phosphatase to remove thiophosphate groups from receptor proteins.

In these experiments inhibition of channel gating was seen in solutions containing 1 mM ATP and $0.12 \mu\text{M Ca}^{2+}$ but without added protein kinase. Assuming that channel inactivation under these conditions results from protein phosphorylation, this observation implies that kinase activity must be associated with the membrane localized within the patch pipette. As calcium/calmodulin-dependent protein kinase II (CaMKII) activated at $0.1 \mu\text{M Ca}^{2+}$ has been isolated from skeletal muscle SR membrane¹⁴⁻¹⁶, we investigated whether a specific peptide inhibitor of CaMKII could prevent channel inactivation (Fig. 3). Patches with high P_o were obtained by pretreating the SR with phosphatase and/or by using a recording solution without added ATP. Addition of 1 mM ATP alone or 1 mM ATP plus $2.0 \mu\text{M}$ control peptide (amino-acid residues 284-302) to the pipette resulted in a decline in average P_o (Fig. 3, top and bottom). The control peptide (284-302) is a truncated form of

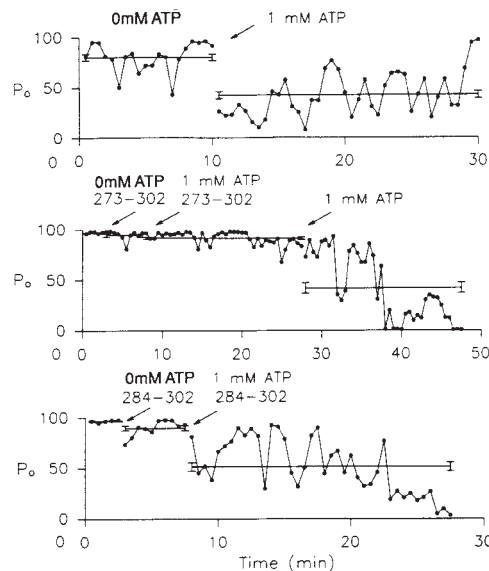


FIG. 3 Peptide inhibitor of multifunctional calcium/calmodulin-dependent protein kinase type II prevents channel inactivation. Time course of changes in P_o following introduction of *a*, 1 mM ATP; *b*, ATP plus the inhibitor peptide of CaMKII (273-302); and *c*, ATP plus a control peptide (284-302). In each case the pipette solution initially contained 0 mM ATP. Both the CaMKII inhibitory peptide and its control were synthesized by the University of Illinois Biotechnology Center Genetic Engineering Facility.

the inhibitory peptide which does not affect CaMKII activity. The inhibitor (273–302) is a 30-amino-acid synthetic peptide that mimics the auto-inhibitory sequence of CaMKII and inhibits CaMKII activity^{17,18}. Addition of 20 μ M CaMKII inhibitory peptide (273–302) prevents the ATP-dependent inactivation of the calcium release channel (Fig. 3, middle). Its effect is reversible; removal of the inhibitory peptide restores ATP sensitivity to the patch. Note that in this experiment, channel P_o varied considerably after the addition of ATP. This was a common finding (never seen when ATP- γ S was the substrate) and may indicate the presence of endogenous phosphatase activity¹⁹. A peptide inhibitor of protein kinase C (corresponding to residues (19–36) did not affect channel gating, suggesting that inactivation results specifically from the activity of CaMKII (Table 1).

Our results provide direct evidence that CaMKII modulates the activity of the SR ryanodine receptor protein/calcium release channel of skeletal muscle when studied in its native lipid environment. An analysis of the published sequence of the ryanodine receptor protein showed no homology with the consensus sequences for the active sites of CaMKII^{20,21}, suggesting

that the channel protein is not capable of autophosphorylation. The calcium release channel is known to be a substrate for endogenous calmodulin-dependent protein kinase^{16,20}, but our results do not allow identification of the protein being phosphorylated in the membrane patch. It has been demonstrated that the cardiac form of the ryanodine receptor protein is more highly phosphorylated by CaMKII than the skeletal form and that phosphorylation increases, rather than decreases, activity of the cardiac receptor in lipid bilayers²². It is possible that phosphorylation of regulatory proteins that coexist with the skeletal release channel in our membrane patches could mediate inactivation^{16,23}. Nonetheless, our results indicate that the calcium release channel from skeletal muscle is closely associated with the kinase and its cofactors. This conclusion follows from the observation that channels isolated in patch pipettes could be modulated by exposure to phosphatase or to ATP and its analogues without addition of exogenous kinase. A similar association between a channel protein and a modulating kinase has recently been proposed for a Ca^{2+} -activated K^+ channel¹⁹. Such associations may prove to be typical for enzymatically modulated ion channels. $\square$

Received 27 April; accepted 4 September 1992.

- Imagawa, T., Smith, J. S., Coronado, R. & Campbell, K. P. *J. Biol. Chem.* **262**, 16636–16643 (1987).
- Hymel, L., Inui, M., Fleischer, S. & Schindler, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 441–445 (1988).
- Lai, F. A., Erickson, H. P., Rousseau, E., Liu, Q.-Y. & Meissner, G. *Nature* **331**, 315–319 (1988).
- Catterall, W. A. *Cell* **64**, 871–874 (1991).
- Smith, J. S. *et al. J. gen. Physiol.* **92**, 1–26 (1988).
- Baylor, S. M., Chandler, W. K. & Marshall, M. W. *J. Physiol.* **344**, 625–666 (1983).
- Schneider, M. F. & Simon, B. *J. Physiol.* **405**, 727–745 (1988).
- Simon, B. J., Klein, M. G. & Schneider, M. F. *J. gen. Physiol.* **97**, 437–471 (1991).
- Kwok, W.-M. & Best, P. M. *Pflügers Arch.* **419**, 166–176 (1991).
- Stein, P. & Palade, P. *Biophys. J.* **54**, 357–363 (1988).
- Hals, G. D., Stein, P. G. & Palade, P. *J. gen. Physiol.* **93**, 385–410 (1989).
- Mori, H., Takisawa, H. & Yamamoto, T. *J. Biochem.* **102**, 263–271 (1987).
- Lüttgau, H. C. & Oetliker, H. *J. Physiol., Lond.* **194**, 51–74 (1968).

- Campbell, K. P. & MacLennan, D. H. *J. Biol. Chem.* **257**, 1238–1246 (1982).
- Tuana, S. T. & MacLennan, D. H. *FEBS Lett.* **235**, 219–223 (1988).
- Chu, A., Sumbilla, C., Inesi, G., Jay, S. D. & Campbell, K. P. *Biochemistry* **29**, 5899–5905 (1990).
- Malinow, R., Schulman, H. & Tsien, R. W. *Science* **245**, 862–865 (1989).
- Wagner, J. A. *et al. Nature* **349**, 793–796 (1991).
- Chung, S., Reinhart, P. H., Martin, B. L., Brautigan, D. & Levitan, I. B. *Science* **253**, 560–562 (1991).
- Takeshima, H. *et al. Nature* **339**, 439–445 (1989).
- Hanks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42–52 (1988).
- Witcher, D. R., Kovacs, R. J., Schulman, H., Cefali, D. C. & Jones, L. R. *J. Biol. Chem.* **17**, 11144–11152 (1991).
- Kim, D. H. & Ikemoto, N. *J. Biol. Chem.* **261**, 11674–11679 (1986).

ACKNOWLEDGEMENTS. We thank Wai-meng Kwok, whose preliminary results demonstrated that adding phosphatase to the bath increased patch activity. This work was supported by grants from the NIH and the Muscular Dystrophy Association (to P.M.B.).

Signal-sequence recognition by an *Escherichia coli* ribonucleoprotein complex

Joen Luijck*, Stephen High, Heather Wood, Angelika Giner, David Tollervey & Bernhard Dobberstein

European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, 6900 Heidelberg, Germany

HYDROPHOBIC signal-sequences direct the transfer of secretory proteins across the inner membrane of prokaryotes and the endoplasmic reticulum membranes of eukaryotes¹. In mammalian cells, signal-sequences are recognized by the 54K protein (M, 54,000) of the signal recognition particle (SRP)^{2,3} which is believed to hold the nascent chain in a translocation-competent conformation until it contacts the endoplasmic reticulum membrane⁴. The SRP consists of a 7S RNA and six different polypeptides. The 7S RNA and the 54K signal-sequence-binding protein (SRP54) of mammalian SRP exhibit strong sequence similarity to the 4.5S RNA and P48 protein (Ffh) of *Escherichia coli*^{5–7} which form a ribonucleoprotein particle. Depletion of 4.5S RNA or overproduction of P48 causes the accumulation of the β -lactamase precursor, although not of other secretory proteins^{8,9}. Whether 4.5S RNA and P48 are part of an SRP-like complex with a role in protein export is controversial. Here we show that the P48/4.5S RNA ribonucleoprotein complex interacts specifically with the signal sequence of a nascent secretory protein and therefore is a signal recognition particle.

SRP54 was identified as the component of the SRP that interacts with the signal sequence of nascent preprolactin and other secretory and membrane proteins by light-induced cross-linking^{2,3}. We used a similar approach to identify proteins in an *E. coli* cell extract that interact with nascent polypeptides. A truncated messenger RNA, encoding the amino-terminal 86 amino-acid residues of preprolactin, was translated in a wheat-germ cell-free translation system, to generate nascent chain-ribosome complexes with an exposed signal sequence. As a control we used an mRNA encoding a mutant preprolactin with three amino-acid substitutions in the hydrophobic core of the signal sequence. This mutated nascent chain is unable to bind SRP54 and is not translocated across the endoplasmic reticulum membrane (not shown). Translation was done in the presence of [³⁵S]methionine to label the polypeptide, and in the presence of ϵ -TDBA 4-(3-trifluoromethyldiazirino)benzoic acid–Lys-transfer RNA, to allow incorporation of photoreactive groups at positions 4 and 9 of the signal sequence. The nascent chain-ribosome complex was then purified and incubated with a crude *E. coli* cell extract.

Ultraviolet irradiation yielded a crosslinked product of M_r 56K (Fig. 1, lane 5) which was not seen in a control, nonirradiated sample (Fig. 1, lane 1). There is a 9K contribution from the nascent chain, and a crosslinked 'partner' of ~48K. The product was not visible when the nascent chain was derived from the mutant preprolactin (Fig. 1, lane 7), indicating that the ~48K protein interacts specifically with functional signal-sequences. The amount of the 56K product increased when the *E. coli* extract was prepared from a strain that overproduces P48 (Fig. 1, lane 6). As 4.5S RNA is synthesized in excess of P48 (ref. 10), it is likely that some of the overproduced P48 interacts with the free 4.5S RNA thereby increasing the amount of P48/4.5S RNA ribonucleoprotein complex. An antiserum directed against P48 immunoprecipitated the crosslinked

* Present address: Department of Molecular Microbiology, Faculty of Biology, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands.

product (Fig. 1, lane 13), whereas a control antiserum did not (Fig. 1, lane 9). The amount of the immunoprecipitated product was increased when the P48-enriched *E. coli* extract was used (Fig. 1, lane 14). Similar results were obtained when the truncated nascent chain, including the N-terminal signal-sequence, was derived from *E. coli* murein lipoprotein (not shown). These results indicate that *E. coli* P48 binds specifically to signal sequences.

A small amount of a 64K product was found on irradiation of nascent mutant preprolactin, incubated in the presence of wild-type and P48-enriched *E. coli* extract (Fig. 1, lanes 7 and 8). This unidentified product was not immunoprecipitated with the anti-P48 serum (Fig. 1, lanes 15 and 16). It is possible that the nonfunctional signal sequence misfolds and binds to a cytosolic chaperone.

When purified canine SRP and *E. coli* cell extract were added simultaneously to nascent preprolactin-ribosome complexes before crosslinking, SRP54 was crosslinked in place of P48 (Fig. 2a, lane 2), indicating that the two proteins bind to the same site in nascent preprolactin.

SRP54 only interacts with nascent proteins¹¹. To test whether this is also true for P48, we disrupted the nascent chain-ribosome complexes by puromycin before adding *E. coli* cell extract. A large reduction in P48 crosslinking was observed (Fig. 2b, lanes 1 and 2). Thus P48 also binds inefficiently, or not at all, to the signal sequences of proteins that have been released from the ribosome.

We then investigated the stability of the interaction between P48 and the preprolactin signal-sequence. P48 in wild-type *E. coli* extract was allowed to bind to nascent chains. Ribosome-nascent chain complexes were treated with high salt and/or puromycin, and then crosslinking was induced. Puromycin treatment virtually abolished P48 crosslinking (Fig. 2c, lanes 3 and 4), suggesting that P48 does not remain associated with polypeptides that have been released from the ribosome. Under these conditions, P48 may be displaced by other cytosolic or membrane components in the *E. coli* extract, although no new crosslinking partners were detected (Fig. 2c, lanes 3 and 4). Alternatively, the nascent chain could undergo conformational changes

when released from the ribosome, thereby losing its affinity for P48. The interaction between P48 and the nascent chain-ribosome complex was substantially resistant to high salt alone (Fig. 2c, lane 2). Similar properties have been reported for SRP54 (ref. 12).

P48 is associated with the 4.5S RNA in *E. coli* lysates^{8,9}, so to determine whether P48 binds to signal sequences as part of a ribonucleoprotein particle, we compared the sucrose gradient mobility of P48, 4.5S RNA and the 56K crosslinked product (see Fig. 1). Nascent preprolactin-ribosome complexes were purified and incubated with wild-type *E. coli* extract. After crosslinking, the products were released from the ribosome by a puromycin/high-salt treatment and separated on a 5–20% sucrose gradient. The fractions were analysed by fluorography, western blotting, and northern blotting (Fig. 3). P48 and 4.5S RNA sedimented primarily in fractions 6 and 7 (Fig. 3b, c), whereas the 56K crosslinked product was most abundant in fraction 7 (Fig. 3a), possibly owing to the contribution of the nascent chain. Under these conditions, most free P48 from a strain that overproduces P48 shows a different distribution. It is found from fraction 3 onwards, probably because of aggregation (not shown), suggesting that all P48 present in a wild-type *E. coli* extract which interacts with nascent preprolactin is present in a complex with 4.5S RNA. To investigate whether P48 also binds to signal sequences as a free protein, an *E. coli* lysate depleted of 4.5S RNA¹³ was used in the crosslinking assay. The efficiency of crosslinking to P48 was greatly reduced (Fig. 2d, lanes 1 and 2), although P48 was present in the lysate as determined by western blotting (not shown). The 4.5S RNA in an *E. coli* extract is thus essential for the signal recognition function of P48.

Taken together, these data indicate that an *E. coli* ribonucleoprotein particle containing P48 and 4.5S RNA interacts specifically with the signal sequence of nascent presecretory proteins, and therefore can be considered as an SRP. This interpretation is supported by recent data (H. D. Bernstein and P. Walter, manuscript in preparation) showing that P48 can be reconstituted into mammalian SRP to replace SRP54 in signal-sequence recognition. Furthermore, depletion of

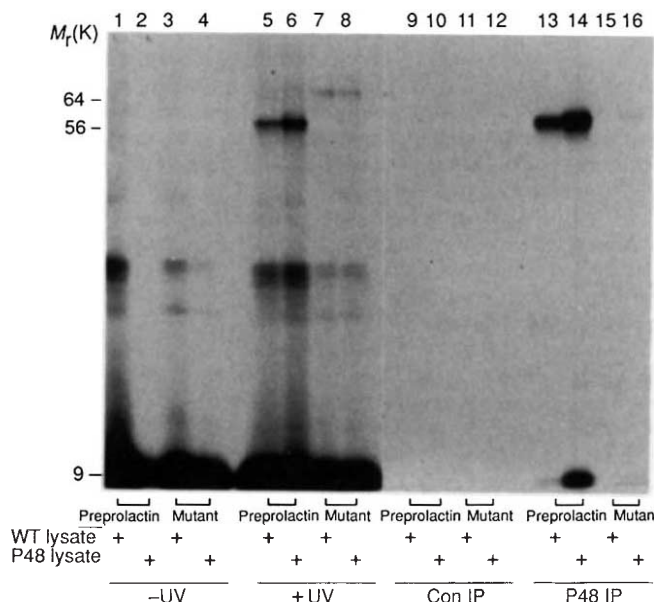
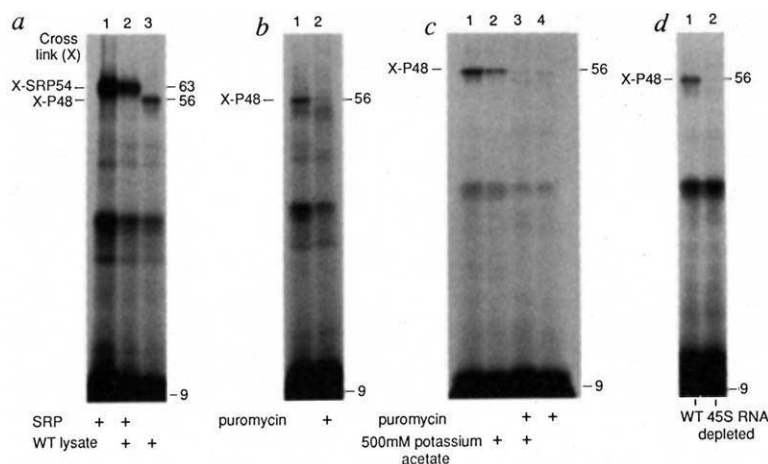


FIG. 1 Specific crosslinking of *E. coli* P48 to nascent preprolactin. A truncated transcript encoding the 86 N-terminal amino-acids of preprolactin or mutant preprolactin was translated in a wheat-germ extract containing [³⁵S]methionine and ϵ -TDBA-Lys-tRNA. The mutant preprolactin carries three amino-acid substitutions in the signal peptide (Leu 15 $\rightarrow$ Pro, Val 19 $\rightarrow$ Glu and Leu 23 $\rightarrow$ Arg) and serves as a control for signal-sequence specificity. Purified ribosome-nascent chain complexes (15 μ l) were incubated with *E. coli* cell extract (10 μ l) in the presence of 2 mM ATP, 40 μ M GTP and an ATP-regenerating system for 5 min at 25 $^{\circ}$ C. The extracts were prepared from a wild-type *E. coli* strain (WT lysate) or from a strain that overproduces P48 (P48 lysate). The samples were chilled on ice for 5 min and subjected to ultraviolet irradiation (+UV; lanes 5–16). Subsequently, the samples were TCA precipitated (lanes 1–8) or immunoprecipitated under denaturing conditions using an antiserum directed against P48 (P48IP, lanes 13–16) or a nonrelated control antiserum (ConIP, lanes 9–12). The samples were analysed on 10–15% SDS-polyacrylamide gels and subjected to fluorography. METHODS. The generation of nascent preprolactin by *in vitro* transcription and translation was as described¹². Ribosome-nascent chain complexes were purified by centrifugation through a 0.5M sucrose cushion¹⁷ and resuspended in 15 μ l buffer (100 mM potassium acetate, 5 mM magnesium acetate, 50 mM HEPES, pH 7.9, and 2 mM cycloheximide) per 20 μ l translation mixture loaded. *E. coli* extracts were prepared from strain MC4100 harbouring either the expression vector pDS12 (WT) or pDS12-48 (P48) which contains the cloned P48 gene under transcriptional control of the T5 promoter⁹. The cells were grown in Luria broth plus 0.4% glucose to an absorbance at 600 nm of 0.3, and P48 overexpression was induced with 1 mM isopropyl- β -D-thiogalactoside (IPTG). The cells were collected after 2 h of induction and lysed by vortexing 40 A_{600} units of cells in 0.5 ml lysis buffer (150 mM KCl, 5 mM MgCl₂, 20 mM Tris-HCl, pH 8.0, 0.1% Triton X-100, 0.1 mM phenylmethylsulphonyl fluoride (PMSF) and 1 mM dithiothreitol) with an equal volume of glass beads (Sigma) for 5 min at 4 $^{\circ}$ C. Immunoprecipitation was done as described¹⁸.

FIG. 2 Characteristics of the interaction between *E. coli* P48 and nascent preprolactin. **a**, SRP inhibits the interaction of P48 with nascent preprolactin. Purified canine SRP (2.5 pmol) (lane 1), 10 μ l wild-type *E. coli* extract (lane 3) or both (lane 2) were added to 15 μ l nascent preprolactin-ribosome complexes before crosslinking. **b**, P48 only interacts with ribosome-bound nascent preprolactin. Purified nascent preprolactin-ribosome complexes were incubated with 500 mM potassium acetate (final) in the presence (lane 2) or absence (lane 1) of 2 mM puromycin for 5 min at 25 °C, before the addition of wild-type *E. coli* lysate and crosslinking. **c**, P48 does not remain bound to nascent preprolactin after release of the P48-nascent chain complex from the ribosome. Nascent preprolactin-ribosome complexes were incubated with wild-type *E. coli* extract as described above and subsequently with H₂O (lane 1), 500 mM potassium acetate (lane 2), 2 mM puromycin (lane 4) or both puromycin and potassium acetate (lane 3) for 5 min at 25 °C followed by crosslinking. **d**, Expression of 4.5S RNA is essential for the interaction of P48 from an *E. coli* extract with nascent preprolactin. Nascent chain-ribosome complexes were incubated with an *E. coli* extract prepared from strain FF283 grown in the presence (lane 1) or absence (lane 2) of IPTG to induce 4.5S RNA expression, and then crosslinking was induced. **METHODS**. The crosslinking assays were as described in the legend to Fig. 1. Canine SRP was purified as described¹⁹. An *E. coli* extract depleted for 4.5S RNA was prepared from strain FF283 in which 4.5S RNA synthesis is



under control of the *Tac* promoter¹³. Cells were grown from fresh plates in M9 medium plus 0.4% glucose for 5 h at 37 °C in the absence of IPTG⁹. After this the culture turbidity of the 4.5S RNA depleted cells was similar to that of cells grown with 1 mM IPTG, indicating that cell death had not occurred.

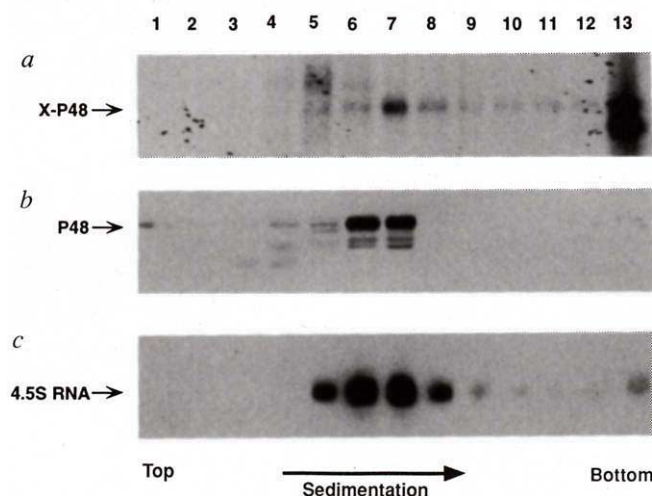


FIG. 3 *E. coli* P48 interacts with nascent preprolactin as part of a ribonucleoprotein particle. Crosslinked complexes were released from the ribosome by addition of 2 mM puromycin and 500 mM potassium acetate followed by incubation for 5 min at 25 °C. The samples were centrifuged for 30 h in a Beckman SW40 rotor at 39,000 r.p.m. and 4 °C through a 5–20% sucrose gradient containing 500 mM KCl, 2 mM MgCl₂, 50 mM tetraethylammonium-HCl, pH 7.5, 0.1% β -mercaptoethanol and 0.1 mM PMSF. Fractions of 1 ml were collected from the top, and portions were analysed by SDS-PAGE and fluorography (**a**), by western blotting using an antiserum directed against P48 (**b**) and by northern blotting using a 4.5S RNA-specific probe (**c**).

METHODS. Crosslinked complexes were prepared as described in Fig. 1 except that emetine was added to 2 mM instead of cycloheximide, to inhibit further chain elongation. Western blotting to detect P48 and northern blotting to detect 4.5S RNA were done essentially as described⁹. Bound antibodies were visualized on the western blot by enhanced chemiluminescence (Amersham).

P48 *in vivo* affects the membrane translocation of several exported proteins (G. J. Phillips and T. Silhavy, manuscript in preparation).

We propose that the role of the *E. coli* SRP is that of a chaperone, that is specific for the signal sequence of nascent preproteins and maintains them in a translocation-competent conformation. Dissociation of the SRP from the signal sequence may be induced by binding of other chaperones either to the nascent chain or to the released polypeptide. Alternatively, the

release of the preprotein from the ribosome may force the dissociation of the SRP. Subsequently, chaperones like SecB and GroEL/ES may maintain the preprotein in an intermediate folding state that is still translocation-competent^{14–16}. Release of the *E. coli* SRP from nascent chains may be mediated by direct interaction with a cognate receptor, similar to the interaction between mammalian SRP and its receptor, the docking protein⁴. Our crosslinking assay should allow the identification of such a receptor. □

Received 1 June; accepted 18 August 1992.

1. von Heijne, G. *Biochim. biophys. Acta* **947**, 307–333 (1988).
2. Krieg, U. C., Walter, P. & Johnson, A. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8604–8608 (1986).
3. Kurzchalia, T. V. *et al. Nature* **320**, 634–636 (1986).
4. Rapoport, T. A. *Trends biochem. Sci.* **15**, 355–358 (1990).
5. Poritz, M. A., Strub, K. & Walter, P. *Cell* **55**, 4–6 (1988).
6. Römisch, K. *et al. Nature* **340**, 478–482 (1989).
7. Bernstein, H. D. *et al. Nature* **340**, 482–486 (1989).
8. Poritz, M. A. *et al. Science* **250**, 1111–1117 (1990).
9. Ribes, V., Römisch, K., Giner, A., Dobberstein, B. & Tollervey, D. *Cell* **63**, 591–600 (1990).
10. Brown, S. *New Biol.* **3**, 430–438 (1991).
11. Wiedmann, M., Kurzchalia, T. V., Bielka, H. & Rapoport, T. A. *J. Cell Biol.* **104**, 201–208 (1987).
12. High, S., Görlich, D., Wiedmann, M., Rapoport, T. A. & Dobberstein, B. *J. Cell Biol.* **113**, 35–44 (1991).

13. Brown, S. & Fournier, M. *J. Cell Biol.* **178**, 533–550 (1984).
14. Randall, L. L. & Hardy, S. J. *Science* **243**, 1156–1159 (1989).
15. Kumamoto, C. A. *Molec. Microbiol.* **5**, 19–22 (1991).
16. Langer, T. *et al. Nature* **356**, 683–689 (1992).
17. High, S., Flint, N. & Dobberstein, B. *J. Cell Biol.* **113**, 25–34 (1991).
18. Römisch, K., Webb, J., Lingelbach, K., Gausepohl, H. & Dobberstein, B. *J. Cell Biol.* **111**, 1793–1802 (1990).
19. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682–691 (1983).

ACKNOWLEDGEMENTS. We thank A. Leyte for comments on the manuscript and T. Ashford for technical assistance, H. Wu and P. Walter for providing antibodies against murein lipoprotein and P48, M. Inouye for the murein lipoprotein clone and T. Silhavy and H. Bernstein for communicating results before publication.

The *E. coli* *ffh* gene is necessary for viability and efficient protein export

Gregory J. Phillips* & Thomas J. Silhavy†

* Department of Biology, The College of William and Mary, Williamsburg, Virginia 23185, USA

† Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

HOMOLOGUES of the gene encoding the 54K (M_r 54,000) subunit of the mammalian signal recognition particle have been identified in different organisms¹⁻⁵. The *Escherichia coli* homologue, termed *ffh* (for fifty-four homologue), specifies a protein (Ffh) that shares many properties with its eukaryotic counterpart, including association with mammalian 7S RNA⁶ and the ability to bind signal sequences specifically^{7,8}. Ffh also associates with *E. coli* 4.5S RNA, showing that it can form a ribonucleoprotein complex in prokaryotes^{6,9,10}. These results are intriguing because extensive genetic and biochemical characterization of *E. coli* failed to identify a signal recognition particle-like mechanism for protein export¹¹. Here we address this issue directly by construction of a strain in which *ffh* expression is arabinose-dependent. Results of depletion experiments indicate that Ffh is important in protein translocation.

The *E. coli* strain WAM113 (for genetic make-up, see Fig. 1) contains a knock-out mutation of the chromosomal copy of *ffh* (and *ffh1::kan*), in combination with a λ prophage carrying *ffh*⁺ under regulation of the *araB* promoter and operator. Preliminary characterization of WAM113 indicated that expression of *ffh* is essential for *E. coli* growth and viability. WAM113 is incapable of forming colonies in the absence of the inducer arabinose (Fig. 2a). Similar results are obtained when cells are plated on either complex L-broth or minimal agar at a variety of growth temperatures. The few colonies that did grow on the plate in Fig. 2a are spontaneously arising arabinose-independent mutants which seem to carry mutations in *araC*.

The arabinose-dependent phenotype of WAM113 provides a means to study the physiological consequences of Ffh depletion

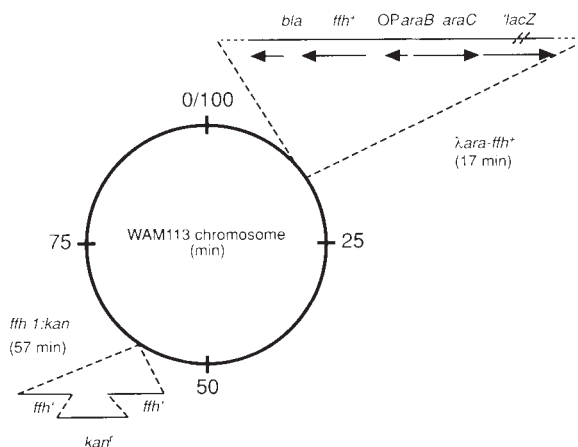
from this *E. coli* strain. A typical growth curve of WAM113 cultured with and without arabinose is shown in Fig. 2b. A difference in growth rates of the two cultures is apparent, particularly at the advanced stages. In addition, we observed that the Ffh-depleted cells exhibited an elongated cell phenotype (Fig. 2c) indicative of a defect in proper cell division. This observation can account for the apparent subtle difference in growth rate of the two cultures as measured by optical density. Indeed, if the experiment is continued by diluting the cultures into fresh medium, growth of the uninduced cells quickly ceases. These results are consistent with the results of Fig. 2a and show that expression of *ffh* is essential for *E. coli* growth.

Of immediate interest was whether cells depleted of the *ffh* gene product are defective in protein secretion, consistent with a role of Ffh as a component of a prokaryotic signal recognition particle (SRP). Cells were grown in the absence of arabinose, as shown in Fig. 2b. At the time indicated, cells in log-phase growth were pulse-labelled with ³⁵S-methionine followed by a chase with unlabelled methionine. Samples of the radiolabelled cells were taken at chase times ranging from 10 s to 5 min. Microscopic inspection of the cultures, grown to the point where they could be radiolabelled, revealed that the cells were elongated (Fig. 2c). Immunoprecipitation was used to monitor the efficiency of signal sequence processing of a variety of exported proteins. All proteins tested exhibited a defect in signal sequence processing (Fig. 3). This defect is most pronounced in the periplasmic proteins alkaline phosphatase, β -lactamase and ribose-binding protein. Even at the longer chase times, precursor forms of these proteins can be detected. In addition, an export defect is seen with the periplasmic maltose-binding protein and the outer membrane proteins, LamB (Fig. 3) and OmpF (not shown).

A number of *E. coli* genes have been identified as components of the bacterial protein export machinery¹², but until now there was no direct evidence that a prokaryotic SRP is important for export of *E. coli* proteins. Reports^{6,9} have focused on disruption of synthesis of wild-type 4.5S RNA, the presumed RNA component of the prokaryotic SRP. Two separate approaches have shown an export defect for β -lactamase only, but this result was apparently caused by a shortage of heat-shock proteins which are thought to be important for maintaining newly synthesized β -lactamase in an export-competent conformation¹³. No

FIG. 1 Chromosomal map of WAM113. Conditional expression of *ffh* is accomplished by a knockout mutation of *ffh* (*ffh1::kan*) in combination with a complementing copy of *ffh* under regulation of the *araB* promoter and operator (*OPara*), carried by λ *ara-ffh*. The chromosomal locations of these genetic elements are shown. Arrows indicate orientation of the genetic elements.

METHODS. The *ffh* knockout mutation was constructed by isolating a 1.2 kb *HincII* restriction fragment containing the gene for kanamycin resistance (*kan*) from pUC4K¹⁷ and inserting it into *HpaI*-digested pBYO3 (ref. 5). This subcloning step results in an insertion/deletion mutation in the open reading frame corresponding to *ffh*, creating the allele *ffh1::kan*. An *EcoRI* to *Sall* fragment from this pBYO3 derivative carrying *ffh1::kan* was subcloned into appropriately digested pMAK705¹⁸. This plasmid is temperature-sensitive for replication and specifically designed for homologous recombination-promoted insertion of DNA sequences into the *E. coli* chromosome. The resulting plasmid, pMAKffh::kan, was used to replace the chromosomal copy of *ffh* with the *ffh1::kan* sequences. Kanamycin-resistant recombinants resulting from *ffh* disruption were obtained only if the recipient carried pBYO3, or a derivative plasmid that included a complementing copy of *ffh*⁺. The *ffh* gene was placed under control of the *araB* operator and promoter by first subcloning a *NruI* to *PstI* fragment carrying a promoterless *ffh* gene from pFfh322 (a pBR322-derivative plasmid constructed for this study) into pBAD18. The resulting construct, pAraffh18, was used as the source for a 3.5 kb *PstI* to *Clal* fragment and cloned into appropriately cut pBR322 DNA, generating pAraffh322. A *PstI* to *HindIII* fragment from this plasmid was used to subclone into pMLB1107, generating pAraffh1107. This plasmid carried *araB-ffh*⁺ along with *araC*⁺, the transcriptional regulator of *araB*, flanked in the appropriate orientations by *bla* and *lacZ*. These sequences



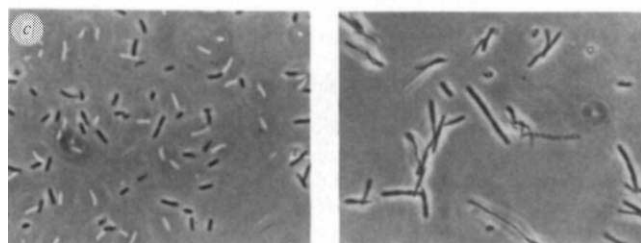
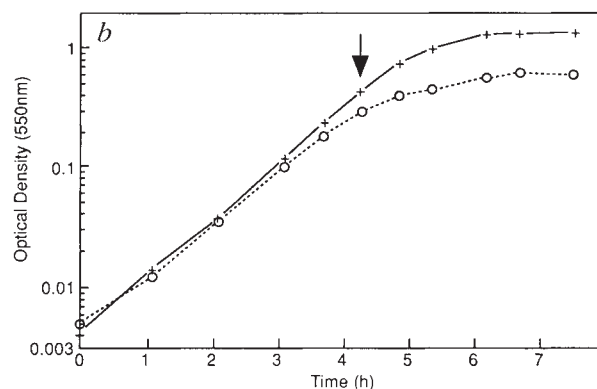
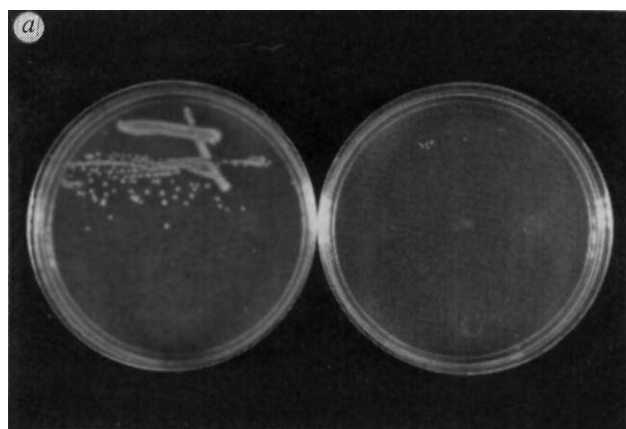


FIG. 2 Growth characteristics of WAM1113, *a*, *ffh* is an essential gene as shown by the arabinose dependence of the strain. Cells were streaked on Luria agar supplemented with (left) and without arabinose. *b*, Growth curve of WAM1113 cultured in MOPS minimal medium²⁰ supplemented either with (+) or without (○) arabinose. Arrow indicates the point at which the cultures were examined microscopically, or radiolabelled as described in Fig. 3. *c*, Morphology of cells grown with (left) or without arabinose. METHODS. Cells were cultured in MOPS minimal medium²⁰ supplemented with amino acids²¹ (excluding cysteine and methionine), 0.001% thiamine and 0.2% fructose. Cells were grown overnight in 5 ml of the same medium supplemented with 0.2% L(-)-arabinose. The next day, the cells were pelleted and washed twice in the same medium lacking a carbon source. Cells (0.01 ml) were used to inoculate 5 ml of the same medium containing either arabinose or 0.2% glucose. Cells were cultured at 37 °C, and optical density measurements made at the times indicated. Cells were visualized by phase contrast microscopy under $\times 330$ magnification using a H2-B Olympus microscope and camera. Photographs were recorded on Kodak T-Max 400 film.

export defects were found with maltose-binding protein, alkaline phosphatase⁶ and ribose-binding protein⁹; proteins that clearly showed defects in signal sequence processing when analysed in cells depleted of Ffh (Fig. 3).

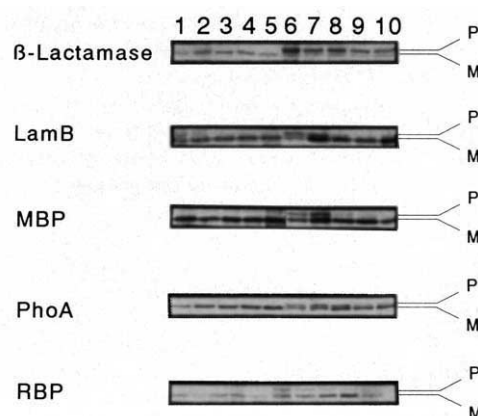
An explanation as to why defects in protein export were not detected by disruption of 4.5S RNA synthesis is that 4.5S RNA may function in two pathways in *E. coli*¹⁰. One pathway involves association with Ffh, shown here to be important for protein export. A second pathway places 4.5S RNA as an integral component of the ribosome, involved in modulating the rate of protein synthesis. The consequences of 4.5S RNA depletion

might be first manifested in decreased rates of protein synthesis, thereby masking any detrimental effects on protein export.

Our results, coupled with the evidence that Ffh and 4.5S RNA interact *in vivo*^{6,9,10}, provide evidence that an SRP-like complex is important for efficient export of *E. coli* proteins. One possible mechanism of action of the prokaryotic SRP is to prevent the premature folding of newly synthesized polypeptides into a conformation that is incompatible with membrane translocation¹⁴. Consistent with this view, we find that the export of some proteins is more adversely affected than others by Ffh depletion. Indeed, the three proteins that exhibit the most dramatic defects

FIG. 3 Signal sequence processing of exported *E. coli* proteins following Ffh depletion. Cells were grown as described in Fig. 2*b*. At the indicated time, cells grown with arabinose (lanes 1–5) or without arabinose (lanes 6–10) were radiolabelled, followed by a chase of 10 s (lanes 1,6), 30 s (lanes 2,7), 1 min (lanes 3,8), 2 min (lanes 4,9) and 5 min (lanes 5,10). 'P' indicates precursor (unprocessed signal sequence) and 'M' indicates mature forms of the exported proteins. Proteins analysed include β -lactamase, LamB, maltose-binding protein (MBP), alkaline phosphatase (PhoA), ribose-binding protein (RBP).

METHODS. Cells were grown as described for Fig. 2*b*. For detection of β -lactamase and PhoA, WAM1113 was transformed with pBR325 or pT7T3PhoA123²², respectively. LamB and MBP production were induced by growing the WAM1113 cultures in 0.2% maltose, which replaced the fructose and glucose supplements. Cultures (2.5 ml) were radiolabelled with 0.02 mCi of Trans-label (ICN Biochemicals) for 30 s. Prewarmed (37 °C) medium (2.5 ml) containing 0.1% unlabelled methionine was then added and 1 ml cells was removed at the indicated times and immediately mixed with 0.05 ml 100% trichloroacetic acid and placed on ice for 30 min. Immunoprecipitations and SDS polyacrylamide gel electrophoresis through 10% gels were performed as described previously²³.



in signal sequence processing, (β -lactamase, alkaline phosphatase and ribose-binding protein) are those that are exported by a mechanism that is independent of the secretion-specific chaperone, SecB^{15,16}. Thus, the role of SRP in *E. coli* seems similar to that played in yeast⁴; both organisms producing some proteins that are more severely affected than others by depletion of the 54K homologue. This result is also consistent with the observation that exported polypeptides have different requirements for antifolding factors, that is, some polypeptides may

be more SRP-dependent than others.

It is unclear why *ffh* was not identified previously, but we note that many of the genetic strategies used were based on LamB or maltose-binding protein and much of the biochemical work has been done with OmpA. Our results show that LamB and maltose-binding protein are only marginally affected by Ffh depletion and this may be true for OmpA as well. Perhaps similar approaches with model proteins such as ribose-binding protein would be rewarding. □

Received 15 June; accepted 11 September 1992.

1. Römisch, K. *et al.* *Nature* **340**, 478–482 (1989).
2. Bernstein, H. D. *et al.* *Nature* **340**, 482–486 (1989).
3. Hann, B. C., Poritz, M. A., & Walter, P. *J. Cell Biol.* **109**, 3223–3230 (1989).
4. Hann, B. C., Walter, P. *Cell* **67**, 131–144 (1991).
5. Byström, A. S., Hjalmarsson, K. J., Wikström, P. M. & Björk, G. R. *EMBO J.* **2**, 899–905 (1983).
6. Ribes, V. *et al.* *Cell* **63**, 591–600 (1990).
7. Bernstein, H. D., Zopf, D., Freymann, D. & Walter, P. (manuscript in preparation).
8. Lührink, J. *et al.* *Nature* **359**, 741–743 (1992).
9. Poritz, M. A. *et al.* *Science* **250**, 1111–1117 (1990).
10. Brown, S. *New Biol.* **3**, 430–438 (1991).
11. Bassford, P. *et al.* *Cell* **65**, 367–368 (1991).
12. Bleker, K., Phillips, G. J. & Silhavy, T. J. *J. Bioenerg. Biomembr.* **22**, 291–310 (1990).
13. Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254–257 (1988).
14. Bernstein, H. D., Rapoport, T. A. & Walter, P. *Cell* **58**, 1017–1019 (1989).

15. Kumamoto, C. A. & Beckwith, J. *J. Bact.* **154**, 253–260 (1983).
16. Kumamoto, C. A. & Beckwith, J. *J. Bact.* **163**, 267–274 (1985).
17. Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
18. Hamilton, C. M. *et al.* *J. Bact.* **171**, 4617–4622 (1989).
19. Silhavy, T. J., Berman, M. L. & Enquist, L. W. *Experiments with Gene Fusions* (Cold Spring Harbor Laboratory, New York, 1984).
20. Neidhardt, F. C., Block, P. L. & Smith, D. F. *J. Bact.* **119**, 736–747 (1974).
21. Davis, R. W., Botstein, D. & Roth, J. *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, New York, 1986).
22. Kohl, J. *et al.* *Nucleic Acids Res.* **18**, 1069 (1990).
23. Phillips, G. J. & Silhavy, T. J. *Nature* **344**, 882–884 (1990).

ACKNOWLEDGEMENTS. We thank J. Beckwith, G. Björk, S. Brown, N. Constantino, D. Court, L. Guzman, S. Kushner and J. Thomas for bacterial strains, technical assistance and helpful suggestions. This work was supported by the NIH (G.J.P. and T.J.S.) and the Howard Hughes Medical Institute (G.J.P.).

Pol of *gag-pol* fusion protein required for encapsidation of viral RNA of yeast L-A virus

Tsutomu Fujimura^{*†}, Juan C. Ribas[†],
Alexandr M. Makhov[‡] & Reed B. Wickner[†]

^{*}Instituto de Microbiología Bioquímica/Departamento de Microbiología y Genética, CSIC/Universidad de Salamanca, Salamanca 37071, Spain

[†]Section on Genetics of Simple Eukaryotes, Laboratory of Biochemical Pharmacology, National Institute of Diabetes, Digestive and Kidney Diseases, Bethesda, Maryland 20892, USA

[‡]National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland 20892, USA

DOUBLE-STRANDED RNA viruses have an RNA-dependent RNA polymerase activity associated with the viral particles which is indispensable for their replication cycle. Using the yeast L-A double-stranded RNA virus we have investigated the mechanism by which the virus encapsidates its genomic RNA and RNA polymerase. The L-A *gag* gene encodes the principal viral coat protein and the overlapping *pol* gene is expressed as a *gag-pol* fusion protein which is formed by a –1 ribosomal frameshift^{1–3}. Here we show that Gag alone is sufficient for virus particle formation, but that it fails to package the viral single-stranded RNA genome. Encapsidation of the viral RNA requires only a part of the Pol region (the N-terminal quarter), which is presumably distinct from the RNA polymerase domain. Given that the Pol region has single-stranded RNA-binding activity, these results are consistent with our L-A virus encapsidation model¹: the Pol region of the fusion protein binds specifically to the viral genome (+) strand, and the N-terminal *gag*-encoded region primes polymerization of Gag to form the capsid, thus ensuring the packaging of both the viral genome and the RNA polymerase.

The L-A double-stranded (ds) RNA virus of *Saccharomyces cerevisiae* contains one L-A dsRNA molecule (4.6 kilobases (kb)) in a 39-nm isometric particle (reviewed in ref. 4). The 5' end of the L-A (+) strand encodes the main coat protein (Gag; *M_r* 76,000) (Fig. 1a). The 3' end gene (*pol*) has the consensus sequence for RNA-dependent RNA polymerases and is expressed as a 170K (*gag-pol*) fusion protein by –1 ribosomal frameshifting^{1–3} like the *gag-pol* fusion proteins of most retroviruses⁵. L-A viral particles have transcriptase activity to synthe-

size (+)single-stranded (ss) RNA^{6,7}. The transcripts are translated into coat proteins^{2,8} and are also the species packaged into new viral particles^{9,10}. We have identified the *cis* packaging site as a 24-nucleotide stem-loop with an A bulge^{11,12}. The encapsidated (+)ssRNA is then replicated to form dsRNA by the viral particle RNA polymerase^{9,10}.

Our model for L-A encapsidation¹ (Fig. 1b) predicts that the Pol region is essential for encapsidation of the viral genome, so here we test the ability of Pol-truncated fusion proteins to package viral RNA into particles. We used a yeast strain with no endogenous L-A virus but carrying two different yeast expression vectors (Fig. 1c). One, the L-A expression plasmid, could constitutively express the L-A genes to provide coat proteins in *trans*¹³; the other, the tester plasmid, contained a small L-A 3' end fragment which included the *cis* packaging site downstream of the GAL-CYC1 promoter¹². Viral particles isolated from the strain having the two plasmids grown on galactose (induced conditions) were purified by differential centrifugation and then on a sucrose gradient. We detected transcripts from the tester plasmid encapsidated in viral particles in the gradient (Fig. 2a), but not when the same cells were grown in glucose (repressed conditions) or when the *cis* packaging site of the tester plasmid was destroyed by a substitution (results not shown)¹². When coat proteins were provided by an endogenous L-A virus, encapsidated tester plasmid transcripts were found in the same position in the gradient (not shown).

When we used a modified expression plasmid (M11A) that had an ochre (TAA) mutation in the *pol* gene at the third codon after the end of the *gag* gene¹³, the transcripts from the tester plasmid were not packaged (Fig. 2b). This could be a result of the inability either of the altered viral proteins to form particles or of the particles made of these proteins to encapsidate the RNA transcripts. Coat proteins in the gradient were detected by polyclonal antiserum against the Gag or Pol regions. The major coat protein was found in the gradient prepared from cells containing the M11A modified expression plasmid (Fig. 3b), and its sedimentation pattern in the gradient was the same as that derived from the intact plasmid (Fig. 3d). As expected, anti-Pol antiserum did not find the *gag-pol* fusion protein in the former gradient (Fig. 3a), but it did detect the fusion protein in the gradient from cells containing the intact plasmid (Fig. 3c). Therefore, the major coat protein by itself is apparently sufficient for particle formation. We also found that proteins expressed from the modified plasmid formed particles that were indistinguishable by electron microscopy from those formed by

proteins from the intact expression plasmid (Fig. 3e and f). As plasmid M11A can still make a truncated fusion protein of 38 amino acids from the *pol* gene which might contribute to particle formation, we changed the frameshifting slippery site of the expression plasmid to abolish frameshifting (1,958-GGGTTT-1,963 changed to 1,958-AGGTTT-1,963; see ref. 3) without modifying the amino-acid sequence of the major coat protein. This mutated expression plasmid gave the same results as M11A (not shown), confirming that the major coat protein is in itself enough for particle formation. Similar results have been reported

for Rous sarcoma¹⁴ and hepatitis B¹⁵ viruses, in which unprocessed Gag or core proteins by themselves can assemble into particles. But particles made of the L-A major coat protein alone fail to encapsidate viral ssRNA having the *cis* packaging site.

We localized the encapsidation domain of Pol by deletion analysis to the N-terminal one quarter (through to residue 213, counting from the frameshifting site) (Fig. 4). Figure 4e shows that these truncated fusion proteins were actually made and incorporated into viral particles, thus eliminating the possibility that the defect of the *NotI*-truncated fusion protein in RNA

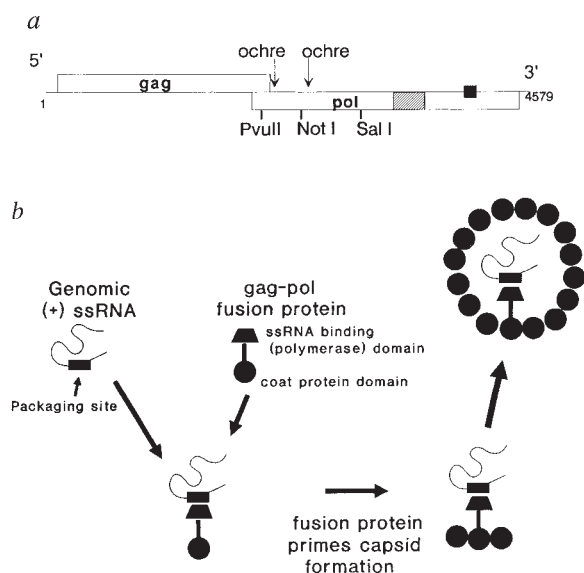


FIG. 1 a, Gene organization of the L-A (+) strand. Pol has single-stranded RNA binding activity and an amino-acid sequence diagnostic of viral RNA-dependent RNA polymerases (hatched region). L-A (+) ssRNA has a *cis* packaging site (filled square) close to the 3' end. The restriction sites and two ochre mutations used to generate truncated *gag-pol* fusion proteins are shown. b, L-A encapsidation model. The 170K fusion protein binds to the *cis* packaging site on the viral (+) strand by its encapsidation domain in the Pol region, and then the N-terminal Gag region primes assembly by homologous interactions with free Gag protein. c, Scheme to assess the ability of L-A coat proteins to encapsidate viral RNA. The expression plasmid (pL2L2) contained all the L-A coding regions in the vector p375, downstream

of a PGK1 promoter¹³. The tester plasmid is the vector pEMBLyex4 with a small L-A fragment (425 bp, from 4,113 to 4,537) containing the *cis* packaging site (4,180 to 4,203) downstream of the GAL-CYC1 hybrid promoter¹². Both plasmids can be maintained in the same cell having *ura3* and *trp1* mutations using their respective selective markers *URA3* and *TRP1*. The galactose-induced transcripts from the tester plasmid containing the *cis* packaging site (shown as a stem-loop with a bulge) can be packaged into viral particles if the coat proteins from the expression plasmid have assembly and encapsidation activities. The encapsidated transcripts can be detected by northern hybridization using a specific probe that recognizes the RNA segment derived from the vector sequence (shown as a wavy line)¹².

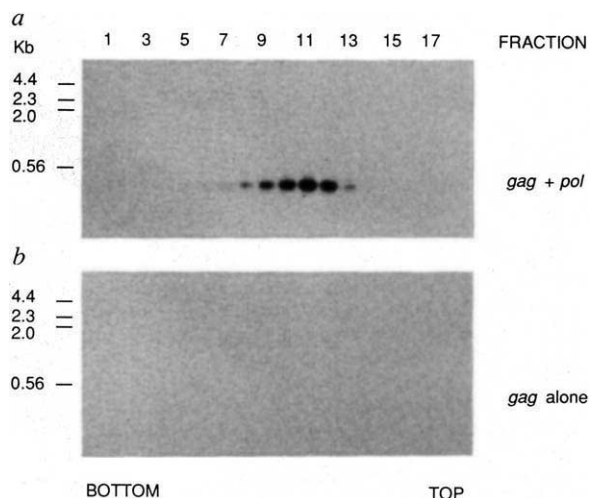


FIG. 2 L-A viral RNA packaging requires the Pol region of the *gag-pol* fusion protein. a, Yeast strain TF395 (*MATa ura3 trp1 leu2 L-A-0*) was transformed with both the expression plasmid and the tester plasmid. Viral particles were prepared from the cells grown on galactose and separated on a sucrose gradient. The tester plasmid transcripts in the gradient were detected by northern hybridization using a specific RNA probe. b, Same as a, except that a modified expression plasmid (M11A) was used. M11A (ref. 13) has an ochre mutation in *pol* at the third codon after the end of *gag*. The migration positions of lambda *HindIII* markers are shown. Fraction 1 corresponds to the bottom of the gradient.

METHODS. Yeast cells were grown for 3 days at 30 °C with galactose in 1 litre of synthetic medium²³ from which uracil and tryptophan had been omitted. Viral particles were prepared by differential centrifugation to remove cell debris and soluble materials²⁴, then particles were precipitated by centrifugation at 10,000g for 10 min after incubation for 30 min with 4.5% PEG 4000 and 0.5 M NaCl. The pellet was dissolved in 2.5 ml of buffer A (50 mM Tris-HCl, pH 7.6, 0.15 M NaCl, 5 mM EDTA and 1 mM DTT). The solution was loaded onto a 36-ml 10–40% (w/v) sucrose gradient in buffer A and centrifuged in a SW27 rotor for 12 h at 17,000 r.p.m. The gradient was fractionated into 22 fractions. RNA was extracted from aliquots (20 µl) of fractions 1 to 18, separated in an agarose gel, denatured in the gel and blotted onto a nylon membrane¹². Transcripts from the tester plasmid were detected by hybridization using ³²P-labelled T3 RNA polymerase transcripts from *HindIII*-cut pRE250 as described¹².

encapsidation is caused by its failure to assemble into particles. This shows that the C-terminal three quarters of the Pol region is unnecessary for the incorporation of the fusion protein into the viral particle and supports the notion¹ that it is directed by homologous interactions of the *gag*-encoded region of the fusion protein with free Gag protein¹. These results also show that the leucine zipper-like sequence (residues 216 to 244) is not necessary for the encapsidation process.

Between the *NotI* site and residue 214 there is a basic amino-

acid sequence (209-SPRRK-213) that could form an Asx-turn, a special type of β -turn^{16,17}. Similar sequences are known to be involved in DNA or RNA binding¹⁵⁻¹⁷, so this sequence may be part of the domain that recognizes and/or binds to the *cis* packaging site. The consensus sequence for RNA-dependent RNA polymerases¹⁸ resides in the C-terminal half of the *pol* gene², far from the region necessary for the RNA encapsidation, so the latter probably forms a distinct functional segment.

Because encapsidation requires particle formation and incor-

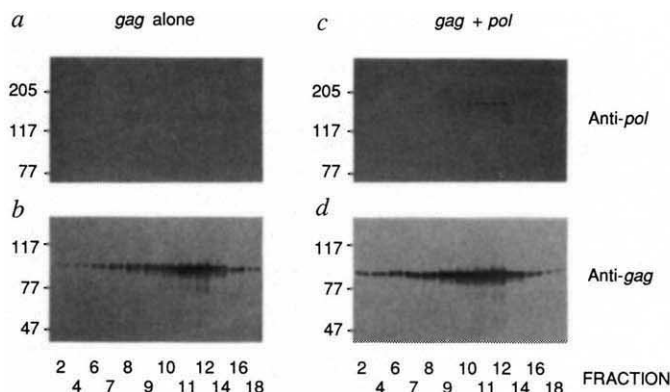


FIG. 3 The major coat protein, Gag, is sufficient for viral particle formation. *a-d*, The presence of viral proteins in the sucrose gradient shown in Fig. 2*b* (prepared from cells expressing only *gag*) was verified using polyclonal anti-Pol (*a*) or anti-Gag (*b*) antiserum. As a control, the gradient shown in Fig. 2*a* (prepared from cells expressing *gag* and *gag-pol*) was also examined with anti-Pol (*c*) or anti-Gag (*d*) antiserum. The migration positions of prestained protein M_r ($\times 10^{-3}$) standards (Bio-Rad) are shown.

METHODS. Proteins in fractions of the sucrose gradients (20 μ l aliquots) were separated in a 7.5% SDS-polyacrylamide gel and electroblotted onto a nitrocellulose sheet as described before¹. The Gag and *gag-pol* fusion proteins were detected with the appropriate antisera using a Protoblot

immunoblotting system (Promega Biotec). The anti-Gag antiserum was raised against Gag protein purified from L-A viral particles as described¹. The anti-Pol antiserum was raised against protein encoded by the *PvuII*-digested fragment of *pol* (amino-acid residues 52 to 860) expressed in *E. coli*. *e* and *f*, Electron micrographs of negatively stained viral particles made of Gag alone (*e*) or made of both Gag and the *gag-pol* fusion protein (*f*). Scale bar, 100 nm. The viral particles from fraction 11 of the sucrose gradients shown in Fig. 2*a* and *b* were absorbed on a glow-discharged thin carbon film and negatively stained as described²⁵; we used 2% uranyl acetate as a stain. Micrographs were recorded at a nominal magnification of 36,000 $\times$ on a Philips 400RT electron microscope.

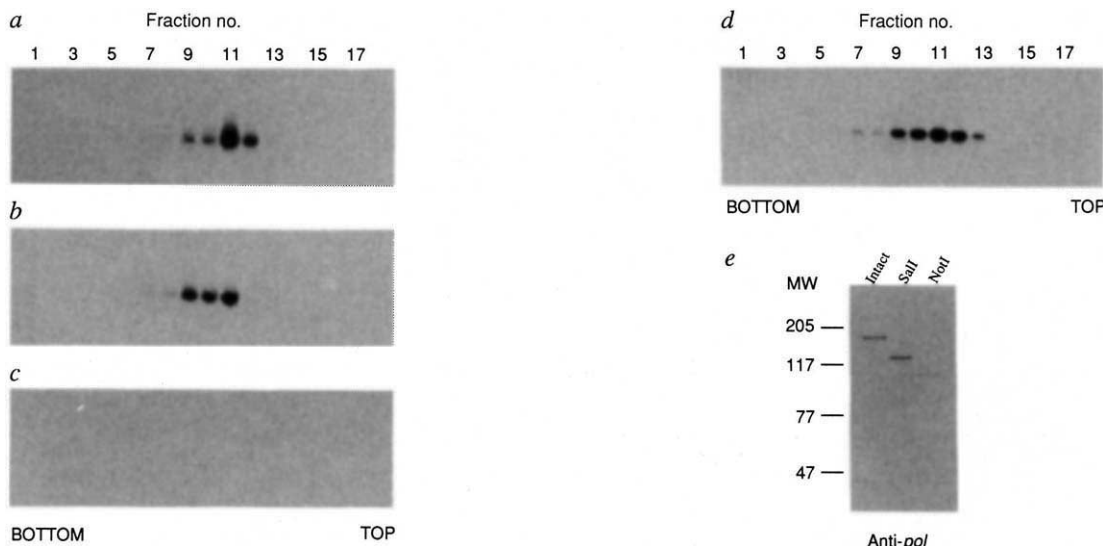


FIG. 4 Deletion analysis of the *gag-pol* fusion protein for its encapsidation activity. *a-d*, Northern blots of sucrose gradients prepared from cells expressing truncated fusion proteins from modified expression plasmids. *a*, The intact fusion protein along with Gag was expressed from pML-9 as control. *b* and *c*, The fusion protein was truncated by deletions from the C terminus to the *SalI* (414 residues of Pol remaining), and *NotI* (205 residues remaining) restriction enzyme sites, respectively. *d*, The fusion protein was truncated by conversion of codon 214 of *pol* to an ochre mutation (pJR10.5), eliminating the leucine zipper-like sequence (residues 216-244). The encapsidated transcripts from the tester plasmid were detected as described in the legend to Fig. 2. The transcripts for the truncated fusion proteins from *SalI*- and *NotI*-deleted expression plasmids

lacked the packaging site, the expression plasmid pML-9 was produced by modification of pL2L2, such that the packaging site on the transcripts from this plasmid was destroyed without modifying the amino-acid sequence in that region. The modified nucleotide sequence was 4,180-TGAGATCGAC-GACCCATCCTTCTG-4,203. pJR10.5 was derived from pML-9. *e*, Incorporation of truncated fusion proteins into viral particles. Protein in fraction number 11 of each sucrose gradient (Fig. 4*a-c*) was separated in a 7.5% SDS-polyacrylamide gel and blotted onto a nitrocellulose sheet as described in the legend to Fig. 3. Truncated fusion proteins were detected with the polyclonal anti-Pol antiserum used for Fig. 3. The migration of prestained protein M_r standards (Bio-Rad) is shown.

poration of the fusion protein into particles, we cannot rule out the direct involvement of the Gag region of the fusion protein in the encapsidation process. We consider, however, that this is unlikely because it is the Pol region, and not the Gag region, that has ssRNA-binding activity¹. At any rate, the direct involvement of the fusion protein in the viral RNA genome encapsidation process ensures that the virus will encapsidate its RNA polymerase and this mechanism provides an advantage for small RNA viruses like L-A with limited genome coding capacity. In hepatitis B virus the polymerase (P) protein is necessary for genomic RNA encapsidation^{19,20}, although the polymerase is not made as a fusion protein with the core protein²¹. If the P gene product interacts with the core protein, this interaction

could substitute for a covalent bond between the Gag and Pol of the L-A fusion protein and thus an encapsidation mechanism similar to L-A's could be applicable. Although the nucleic-acid binding (NC) protein or its unprocessed Gag precursor is believed to be responsible for retroviral genome packaging²², the encapsidation mechanism of L-A resembles that of retroviruses in the way in which Pol is made and encapsidated.

Interactions between the Gag and Pol regions of the L-A fusion protein might also prevent premature activation of the RNA polymerase before completion of packaging, or stimulate interaction of the Gag domain with free Gag protein when Pol has found a (+) strand to bind. Such mechanisms would promote the efficiency of intact viral particle formation. □

Received 25 June; accepted 8 September 1992.

1. Fujimura, T. & Wickner, R. B. *Cell* **55**, 663–671 (1988).
2. Icho, T. & Wickner, R. B. *J. biol. Chem.* **264**, 6716–6723 (1989).
3. Dinman, J. D., Icho, T. & Wickner, R. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 174–178 (1991).
4. Wickner, R. B. in *The Molecular Biology of the Yeast Saccharomyces* Vol. 1 (eds Broach, J. R., Jones, E. W. & Pringle, J. R.) 263–296 (Cold Spring Harbor Laboratory Press, New York, 1991).
5. Jacks, T., Madhani, H. D., Masia, F. R. & Varmus, H. E. *Cell* **55**, 447–458 (1988).
6. Herring, A. J. & Bevan, E. A. *Nature* **268**, 464–466 (1977).
7. Welsh, D. J. & Leibowitz, M. J. *Nucleic Acids Res.* **8**, 2365–2375 (1980).
8. Hopper, J. E., Bostian, K. A., Rowe, L. B. & Tipper, D. J. *J. biol. Chem.* **252**, 9010–9017 (1977).
9. Fujimura, T., Esteban, R. & Wickner, R. B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4433–4437 (1986).
10. Fujimura, T. & Wickner, R. B. *Molec. cell. Biol.* **7**, 420–426 (1987).
11. Esteban, R., Fujimura, T. & Wickner, R. B. *EMBO J.* **8**, 947–954 (1989).
12. Fujimura, T., Esteban, R., Esteban, L. M. & Wickner, R. B. *Cell* **62**, 819–828 (1990).
13. Wickner, R. B., Icho, T., Fujimura, T. & Widner, W. R. *J. Virol.* **65**, 155–161 (1991).

14. Oertle, S. & Spahr, P. *J. Virol.* **64**, 5757–5763 (1990).
15. Birnbaum, F. & Nassal, M. *J. Virol.* **64**, 3319–3330 (1990).
16. Suzuki, M. *EMBO J.* **8**, 797–804 (1989).
17. Churchill, M. E. A. & Travers, A. A. *Trends Biochem. Sci.* **16**, 92–97 (1991).
18. Kamer, G. & Argos, P. *Nucleic Acids Res.* **12**, 7269–7282 (1984).
19. Hirsch, R. C., Lavine, J. E., Chang, L.-J., Varmus, H. E. & Ganem, D. *Nature* **344**, 552–555 (1990).
20. Bartenschlager, R., Junker-Niepmann, M. & Shaller, H. *J. Virol.* **64**, 5324–5332 (1990).
21. Schlicht, H.-J., Radziwill, G. & Shaller, H. *Cell* **56**, 85–92 (1989).
22. Coffin, J. M. in *Fundamental Virology* 2nd edn (eds Fields, B. N. et al.) 645–708 (Raven, New York, 1991).
23. Wickner, R. B. *Cell* **21**, 217–226 (1980).
24. Fujimura, T. & Wickner, R. B. *J. biol. Chem.* **267**, 2708–2713 (1992).
25. Valentine, R. C. & Green, N. M. *J. molec. Biol.* **27**, 615–617 (1967).

ACKNOWLEDGEMENTS. We thank R. Esteban for preparing and critically reading the manuscript.

Lipid modification at the N terminus of photoreceptor G-protein α -subunit

Koichi Kokame*, Yoshitaka Fukada*†, Tōru Yoshizawa‡, Toshifumi Takao§ & Yasutsugu Shimonishi§

* Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606-01, Japan

† Department of Applied Physics and Chemistry, The University of Electro-Communications, Tokyo 182, Japan

§ Institute for Protein Research, Osaka University, Osaka 565, Japan

MYRISTATE is a fatty acid (fourteen-carbon chain with no double bonds, C14:0) linked to the amino-terminal glycine of several proteins^{1–7}, including α -subunits of heterotrimeric ($\alpha/\beta\gamma$) G proteins^{8,9}. We report here a novel modification at the N terminus of the α -subunit of the photoreceptor G protein transducin, $T\alpha$, with heterogeneous fatty acids composed of laurate (C12:0), unsaturated C14:2 and C14:1 fatty acids, and a small amount (~5%) of myristate. Both the GTPase activity of $T\alpha/T\beta\gamma$ and the $T\beta\gamma$ -dependent ADP-ribosylation of $T\alpha$ catalysed by pertussis toxin were inhibited by the lauroylated and myristoylated N-terminal peptide of $T\alpha$. The myristoylated peptide gave 50% inhibition at a 3.5 to ~4.5-fold lower concentration than the lauroylated peptide in each assay, indicating that the strength of the interaction between $T\alpha$ and $T\beta\gamma$ is altered by heterogeneous fatty acids linked to $T\alpha$. This suggests that a looser subunit interaction in transducin which is due to an abundance of N-linked fatty acids other than myristate would favour the rapid turnover and catalysis essential for the visual excitation in photoreceptor cells.

The N-terminal sequence of $T\alpha$ ^{10–12} resembles that of myristoylated $G_{i\alpha}$ or $G_{o\alpha}$ (NH₂-Met-Gly-X-X-Ser)^{8,9,13}, whose initiator methionine is removed, exposing an α -amino group of the glycine which is N-myristoylated. $T\alpha$ can be myristoylated

when expressed transiently in COS cells¹⁴, although myristate has not been detected in photoreceptor $T\alpha$ ⁹. On an electrospray ionization mass spectrum (Fig. 1), bovine $T\alpha$ gave an M_r of 40,049 (± 7), which was 215 higher than that deduced from the complementary DNA^{10–12} (39,834; Gly 2–Phe 350). Because $T\alpha$ had a blocked N terminus¹⁵, it was suggested that the N terminus would be modified with unknown group(s), whose mass is close to that (211.2) of the myristoyl moiety.

To investigate the N-terminal structure, $T\alpha$ was digested with a lysyl endopeptidase, and the proteolytic fragments isolated by reversed-phase high-performance liquid chromatography (HPLC; Fig. 2a) were subjected to amino-acid analyses. Unexpectedly, the amino-acid compositions of three peak fractions (N1–N3, Fig. 2a) coincided well with a calculated value of the N-terminal nonapeptide of $T\alpha$, Gly 2–Lys 10. It was concluded that these peptides were derived from the blocked N-terminal part of $T\alpha$, because all the peptides in the three fractions were refractory to Edman degradation, in the same way as intact $T\alpha$. Fast-atom bombardment (FAB) mass spectrometry demonstrated that fraction N1 consisted of two peptides with masses of 1,000.4 (N1a) and 1,024.4 (N1b), which were separated on a phenyl column (Fig. 2a, inset). The observed mass of a peptide in fraction N2 or N3 was 1,026.4 or 1,028.4, respectively. These values, together with a mass (818.4) calculated from the nonapeptide sequence, predicted four kinds of fatty acids, C12:0, C14:2, C14:1 and C14:0 linked by an amide bond to the N-terminal glycine of the nonapeptide N1a, N1b, N2 and N3, respectively (Table 1).

To verify the structure, we synthesized the N-lauroylated or N-myristoylated nonapeptide of $T\alpha$ (designated C12:0- or C14:0-peptide, respectively). Chromatographic analyses (Fig. 2b, c) led us to conclude that N1a and N3 are identical with C12:0- and C14:0-peptide, respectively. The low content of myristoylated $T\alpha$ (N3, 5.2%; Table 1) agrees with the previous study⁹ which failed to reveal myristoylation of $T\alpha$. For identification of positions of the double bonds in C14:2 and C14:1, peptides N1b and N2 were completely digested with carboxypeptidase Y. Negative-ion FAB tandem mass spectrometric analyses of the acyl-glycine in the digest (C14:2-Gly and C14:1-Gly derived from N1b and N2, respectively),

† To whom correspondence should be addressed.

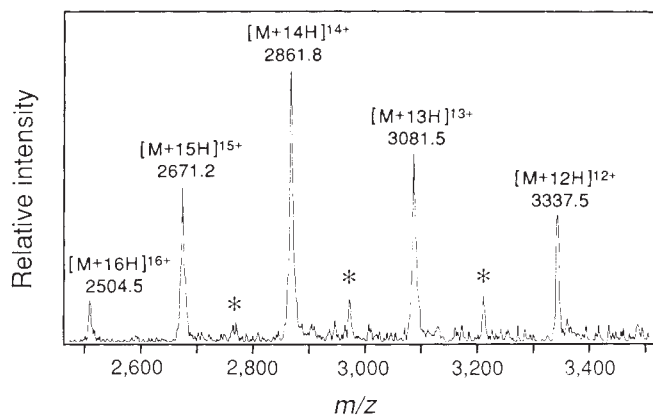


FIG. 1 The electrospray ionization (ESI) mass spectrum of intact $T\alpha$. The ESI mass spectrum was measured by JMS-SX102 mass spectrometer (JEOL) equipped with an ESI ion source (Analytica of Branford, Connecticut). $T\alpha$ was purified from freshly dissected bovine retinas as described previously²⁸. Purified $T\alpha$ (1 nmol) was dissolved in 30 μ l H_2O :acetic acid:methanol (5:1:4) and infused into the ion source at a flow rate of 1 μ l min^{-1} . From each m/z value of the multiply charged ion signals ($[M+nH]^{n+}$), the mass (M) of $T\alpha$ was calculated to be 40,038, 40,047, 40,051, 40,053 or 40,056 ($n=12-16$, respectively). On the average, the M , was 40,049 (± 7 s.d.). The signals marked by asterisks arose probably from $T\alpha$ dimerized during the measurement.

resulted in specific fragmentations¹⁶ which enabled us to assign the locations of double bonds, $\Delta^{5,8}$ and Δ^5 in C14:2 and C14:1, respectively.

Both the synthetic C12:0- and the C14:0-peptide inhibited GTPase activity of transducin ($T\alpha/T\beta\gamma$) in the presence of light-activated receptor, metarhodopsin II (Fig. 3a). This is consistent with observations that the extreme N-terminal regions of many kinds of G protein α subunits are required for interaction with the $\beta\gamma$ subunit¹⁷⁻²⁰. To confirm that the inhibition can be ascribed to competitive effects of the peptides on the interaction between $T\alpha$ and $T\beta\gamma$, we examined the effects of the peptides on $T\beta\gamma$ -dependent ADP-ribosylation of $T\alpha$ cata-

lysed by pertussis toxin in the absence of metarhodopsin II. The reaction, which has been used to detect the interaction between $T\alpha$ and $T\beta\gamma$ ²¹, was accelerated 20-fold by the addition of $T\beta\gamma$ under the conditions (compare $\blacktriangledown$ with ∇ in Fig. 3b). This increment in initial rate of the reaction was completely reduced to the $T\beta\gamma$ -independent level by the addition of increasing amounts of C12:0- and C14:0-peptide (Fig. 3b). But no effect was observed on either the GTPase activity or the ADP-ribosylation when the N-terminal fatty acid of the effective peptide was replaced by an acetyl group (C2-peptide in Fig. 3a, b). These results indicate that the N-terminal acyl chain of $T\alpha$ is important for the functional interaction with $T\beta\gamma$. This is compatible with

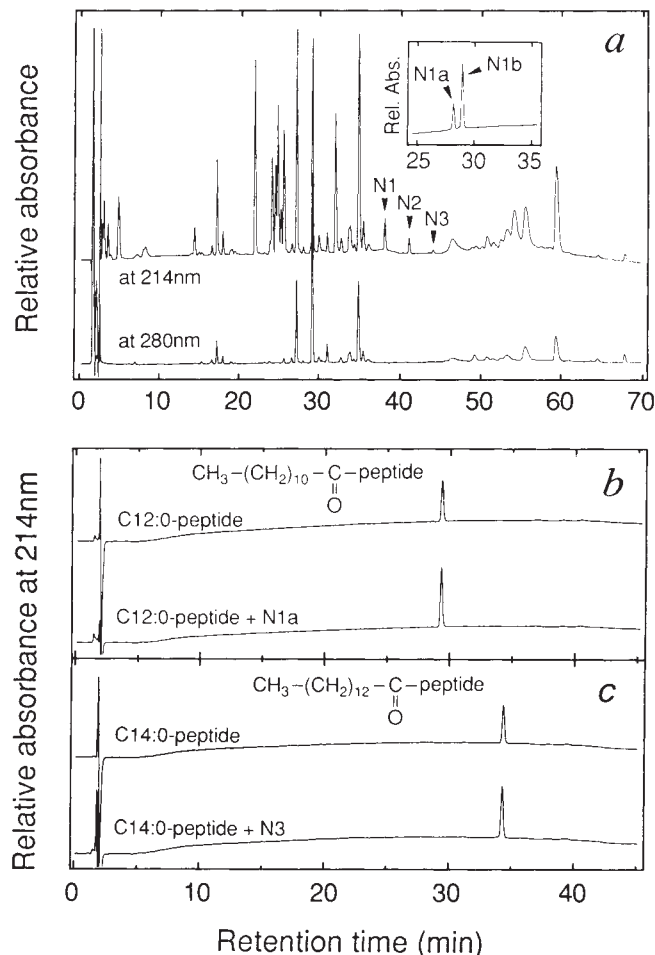


FIG. 2 Isolation and identification of N-terminal proteolytic fragments of $T\alpha$. a, Purified $T\alpha$ (13.0 nmol) was digested by incubation at 37 $^{\circ}C$ for 12 h with 8 μ g *Achromobacter lyticus* protease I (Wako) in 1 ml 12.5 mM NH_4HCO_3 (pH 7.8). The digest was loaded onto a reversed-phase C_{18} -column (Develosil ODS-T-5; 4.6 $\times$ 150 mm, Nomura Chemical) equipped with a HPLC system (Model LC204, Waters). Fragments were eluted with a linear gradient of 10–80% acetonitrile (1% min^{-1}) in 0.1% trifluoroacetic acid at a flow rate of 1 ml min^{-1} . The absorbances of the eluate at 214 nm and 280 nm were simultaneously monitored. Inset, fraction N1 was further chromatographed on a phenyl-column (Cosmosil 5PhT-300; 4.6 $\times$ 150 mm, Nacalai Tesque) equipped with the HPLC system. Two peptides, N1a and N1b, were separated from each other under the same elution conditions as described above. A part of the elution profile monitored at 214 nm is shown. b, c, The synthetic C12:0-peptide (b, upper trace) or C14:0-peptide (c, upper trace) was loaded onto the Develosil ODS-T-5 column, and eluted with a linear gradient of 20–80% acetonitrile (1% min^{-1}) in 0.1% trifluoroacetic acid at a flow rate of 1 ml min^{-1} . Under these conditions, an equimolar mixture of C12:0-peptide and N1a (b, lower trace), or C14:0-peptide and N3 (c, lower trace) was eluted from the column in a single peak. On another column (Cosmosil 5PhT-300), each mixture also gave a single peak (data not shown), indicating that N1a and N3 are identical with C12:0- and C14:0-peptide, respectively. Synthesis of acylpeptides: The N-terminal nonapeptide of $T\alpha$ (Gly-Ala-Gly-Ala-Ser-Ala-Glu-Lys) was synthesized by an automated peptide synthesizer (9050 PepSynthesizer, MilliGen). Before deprotection, the peptide (18 μ mol) attached to the resin was reacted with myristoyl or lauroyl succinimide (72 μ mol) in 1 ml of *N,N*-dimethylformamide at room temperature for 9 h. Alternatively, the parent peptide (18 μ mol) was acetylated by incubating it in 800 μ l of 10% acetic anhydride (in *N,N*-dimethylformamide) at room temperature for 6 h. The acylpeptides thus produced were cleaved from the resin and deprotected to obtain lauroylated, myristoylated or acetylated peptide (C14:0-, C12:0- or C2-peptide, respectively), which was purified by the reversed-phase HPLC. In FAB mass spectrometric analyses, the observed mass of purified C14:0-peptide (1,028.4), C12:0-peptide (1,000.4) or C2-peptide (860.4) agreed well with the theoretical value (1,028.6, 1,000.5 or 860.4, respectively).

TABLE 1 Amino-acid compositions and mass values of N1a, N1b, N2 and N3

	N1a	N1b	N2	N3	Calculated		N1a	N1b	N2	N3	Calculated
Amino-acid composition						Relative					
Ala	3.00	2.94	2.97	2.74	3	quantity (%)	22.6	35.4	29.5	5.2	
Glx	2	2	2	2	2	Observed M_r	1,000.4	1,024.4	1,026.4	1,028.4	
Gly	2.07	2.02	1.96	2.00	2	Linked fatty acid	C12:0	C14:2	C14:1	C14:0	
Lys	0.95	0.94	0.94	0.96	1			(Δ^{58})	(Δ^5)		
Ser	0.96	0.93	0.81	0.85	1						

Amino-acid compositions of the four fractions were analysed in an amino-acid analyser (L8500, Hitachi), and normalized to the quantities (two residues) of Glx (sum of Gln and Glu). All the values agreed well with the theoretical values of the N-terminal nonapeptide of rod $T\alpha^{10-12}$ (calculated), and distinct from those of cone $T\alpha^{27}$ (Ala, 2; Asp, 1; Glu, 1; Gly, 2; Lys, 1; Ser, 2). In the amino-acid analyses, the amount of the peptide was calculated to be 2.94 nmol (N1a), 4.60 nmol (N1b), 3.84 nmol (N2) or 0.68 nmol (N3). The relative quantity of each peptide is shown as per cent of total amount of $T\alpha$ (13.0 nmol) subjected to this experiment. The relative quantity (%) did not vary significantly among different batches of purified $T\alpha$. The mass of each peptide was measured by FAB mass spectrometry using a double-focusing mass spectrometer (JMS-HX100, JEOL). A fatty acid linked to each peptide was predicted on the basis of a mass value obtained by subtracting the theoretical mass of the nonapeptide (818.4) from the observed mass. The peptides N1b and N2 (~1 nmol each) were digested with 1 μ g carboxypeptidase Y (Takara) in 20 μ l 10 mM ammonium acetate buffer (pH 5.5) at 25 °C for 16 h. Then the positions of the double bonds in C14:2 and C14:1 were determined by using a four-sector tandem mass spectrometer (JMS-HX/HX110A, JEOL).

myristoylation of $Go\alpha$ being indispensable for its functional coupling with the $\beta\gamma$ subunit²².

There is a remarkable difference between the C12:0- and the C14:0-peptide in the concentration required for half maximal inhibition (I_{50}) of the GTPase activity and the ADP-ribosylation reaction (Fig. 3a, b). Such a difference in I_{50} value between C12:0- and C14:0-peptide (740 μ M and 210 μ M for the GTPase

or 680 μ M and 150 μ M for the ADP-ribosylation, respectively) indicates that $T\beta\gamma$ interacts more loosely with *N*-lauroylated $T\alpha$ than with *N*-myristoylated $T\alpha$, the latter of which is a minor subspecies (5.2%) of total $T\alpha$. A random peptide (Gly-Lys-Glu-Glu-Ala-Ser-Ala-Gly-Ala) showed a similar inhibitory effect on the ADP-ribosylation when the peptide was *N*-lauroylated (I_{50} , 560 μ M), *N*-myristoylated (I_{50} , 210 μ M) or *N*-acetylated (no

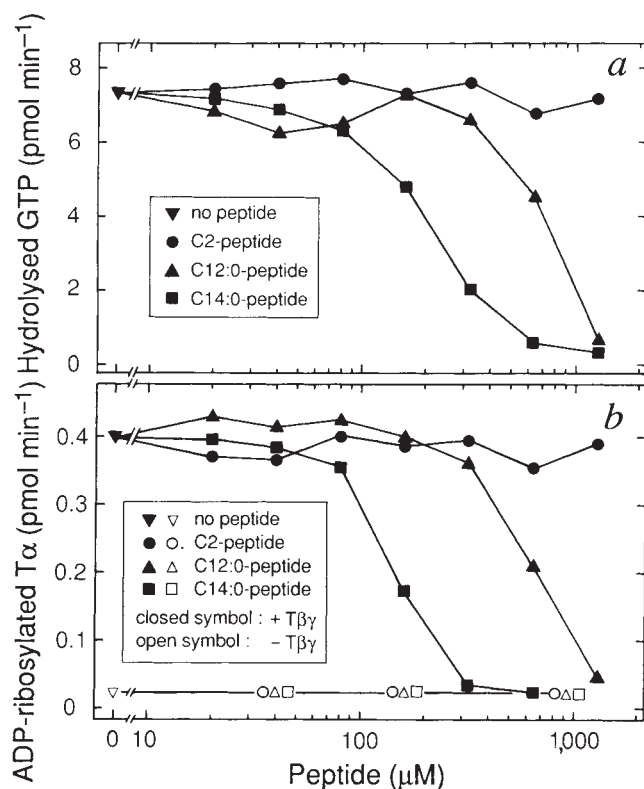


FIG. 3 Effects of synthetic peptides on GTPase activity of transducin (a) and ADP-ribosylation of $T\alpha$ by pertussis toxin (b). a, Rhodopsin reconstituted in phosphatidylcholine liposomes²⁸ was irradiated with an orange light (>540 nm) for 1 min at 4 °C to convert rhodopsin into metarhodopsin II. The reaction mixtures (50 μ l) were composed of metarhodopsin II in liposomes (0.1 μ M), $T\alpha$ (0.2 μ M), $T\beta\gamma$ (0.2 μ M), [γ -³²P]GTP (10 μ M, 157 mCi mmol⁻¹), soybean trypsin inhibitor (0.8 mg ml⁻¹), and various concentrations of synthetic peptides in buffer A (10 mM 3-(*N*-morpholino)propanesulphonic acid-NaOH, 100 mM NaCl, 2 mM MgCl₂, 1 mM dithiothreitol, 0.1 mM phenylmethylsulphonyl fluoride, 4 μ g ml⁻¹ leupeptin, and 50 kallikrein inhibitor units ml⁻¹ aprotinin; pH 7.5). After 20 min incubation at 25 °C, 250 μ l 5% charcoal (in 50 mM Na-phosphate buffer, pH 6.5) was added to the reaction mixture, and centrifuged to get a supernatant containing ³²Pi produced by the hydrolysis of [γ -³²P]GTP. GTPase activity was estimated from the radioactivity in the supernatant, and plotted against the concentration of added peptide. The data shown are representative results of four independent experiments. b, Pertussis toxin (Kaken-Seiyaku) was preactivated at 37 °C for 10 min in buffer B (72 mM *N*-(2-hydroxyethyl)piperazine-*N'*-(2-ethanesulphonic acid)-NaOH, 180 mM NaCl, 18 mM EDTA, 10 mM dithiothreitol, 3.6 mM ATP, 0.7 mM GDP, 36 mM thymidine; pH 6.6). ADP-ribosylation reaction of $T\alpha$ was started by the addition of buffer B (10 μ l) containing both the preactivated pertussis toxin (25 μ g ml⁻¹) and [α -³²P]NAD (50 μ M, 800 mCi mmol⁻¹), to the mixture (40 μ l; in buffer A) composed of $T\alpha$ (1.25 μ M), $T\beta\gamma$ (0.625 μ M), soybean trypsin inhibitor (1.6 mg ml⁻¹), 0.0125% Lubrol PX, and various concentrations of synthetic peptides. Lubrol PX (final concentration of 0.01%) was included in the reaction mixture to eliminate a possible formation of acylpeptide micelles. After 10, 20 or 30 min of incubation at 37 °C, 800 μ l ice-cold 100 mM Tris-HCl (pH 7.5) containing 1 mM MgCl₂ was added to the reaction mixture, and immediately [³²P]ADP-ribosylated $T\alpha$ was isolated from free [³²P]NAD according to the filtration method²⁸. Under these conditions, the amount of ADP-ribosylated $T\alpha$, which was estimated from the radioactivity of the filter, linearly increased with incubation time (<30 min). The initial rate of the reaction was plotted against the concentration of added peptide (closed symbols). ADP-ribosylation of $T\alpha$ in the absence of $T\beta\gamma$ was done similarly (open symbols). The data shown are representative results of three independent experiments. The γ -subunit of $T\beta\gamma$ used in these experiments is farnesylated and carboxyl methylated at the C-terminal cysteine residue²⁹.

effect; <1.3 mM). Therefore the fatty acid rather than the amino-acid sequence would play a major role in the specific subunit interaction. We suggest that, compared with other *N*-myristoylated G proteins, $T\alpha$ in its lauroyl form would have a looser interaction with $T\beta\gamma$. This would favour the rapid association-dissociation cycles catalysed by metarhodopsin II within a photoresponse time of vision²³.

Myristate has also been postulated to act as a membrane anchor of $G\alpha$ or $G\beta$ which dissociates from the $\beta\gamma$ subunit after activation by receptor and GTP. Unlike the other G proteins, however, transducin is released from membranes *in vitro* when activated by metarhodopsin II and GTP under hypotonic conditions. This unaccounted phenomenon can be explained partly by the abundance of $T\alpha$ modified with C12:0, C14:2 and C14:1 (~90% in total, Table 1), all of which are less hydrophobic than C14:0 (myristate)²⁴. This variation of the modification, probably providing weak interaction with a lipid bilayer, should enable GTP-bound $T\alpha$ to skate along the narrow interface between disk membranes and to interact rapidly with its target enzyme, cyclic GMP-phosphodiesterase.

To our knowledge, this is the first demonstration that the N-terminal glycine is modified with multiple fatty acids including unsaturated ones. It will be interesting to see whether the known myristoyl-CoA, protein *N*-myristoyltransferase (NMT), is responsible for the heterogeneous modifications or whether other transferases exist. Cell-specific difference in acyl-CoA composition may play a dominant role in determining what kind of fatty acid is transferred to nascent polypeptides: CoA thioesters of not only C14:0 but also C14:1 and C12:0 actually serve as substrates for the NMT from *Saccharomyces cerevisiae*^{24,25}, though the N-terminal octapeptide of $T\alpha$ is not myristoylated by the yeast NMT but by rat NMT²⁶. One of the keys to understanding the physiological importance of the N-terminal heterogeneity could be the enzymatic control of the modification. □

Crystal structure of a streptococcal protein G domain bound to an Fab fragment

Jeremy P. Derrick* & Dale B. Wigley†

* Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

† Department of Chemistry, University of York, York YO1 5DD, UK

PROTEIN G is a cell-surface protein from *Streptococcus*^{1,2} which binds to IgG molecules from a wide range of species with an affinity comparable to that of antigen^{3,4}. The high affinity of protein G for the Fab portion of IgG poses a particular challenge in molecular recognition, given the variability of heavy chain subclass, light chain type and complementarity-determining regions. Here we report the crystal structure of a complex between a protein G domain and an immunoglobulin Fab fragment. An outer β -strand in the protein G domain forms an antiparallel interaction with the last β -strand in the constant heavy chain domain of the immunoglobulin, thus extending the β -sheet into the protein G. The interaction between secondary structural elements in Fab and protein G provides an ingenious solution to the problem of maintaining a high affinity for many different IgG molecules. The structure also contrasts with Fab-antigen complexes, in which all contacts with antigen are mediated by the variable regions of the antibody, and to our knowledge provides the first details of interaction of the constant regions of Fab with another protein.

To investigate the molecular basis for the wide specificity of protein G for IgG, we prepared crystals of the complex between the third IgG-binding domain of protein G and an Fab fragment from mouse IgG1 (ref. 5). The structure was solved by molecular replacement using the model for the HyHel-10 Fab/lysozyme complex (Brookhaven; accession number 3HFM) with the

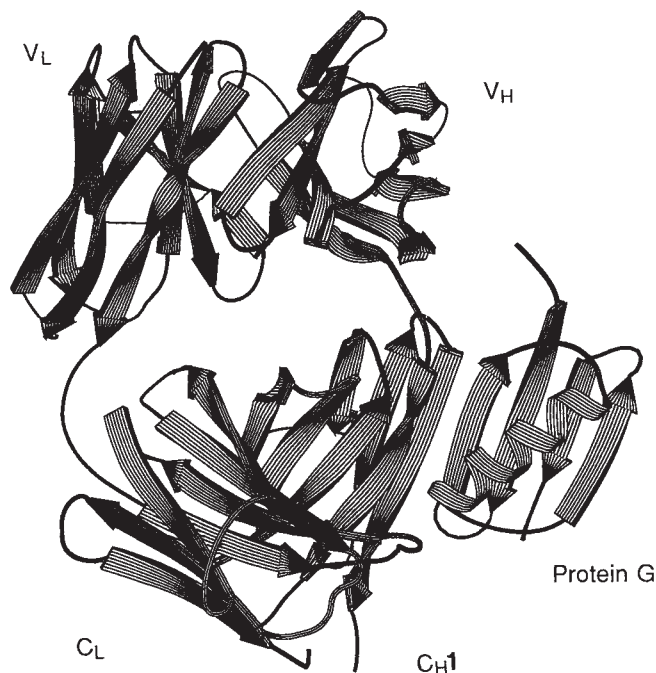


FIG. 1 Overall structure of the Fab/protein G complex. The variable and constant domains of both the heavy and light chains of the Fab are labelled. Protein G interacts with the constant domain of the heavy chain (CH1).

Received 14 April; accepted 24 August 1992.

1. Carr, S. A., Biemann, K., Shoji, S., Parmelee, D. C. & Titani, K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6128-6131 (1982).
2. Aitken, A. et al. *FEBS Lett.* **150**, 314-318 (1982).
3. Henderson, L. E., Krutzsch, H. C. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 339-343 (1983).
4. Ozols, J., Carr, S. A. & Strittmatter, P. *J. biol. Chem.* **259**, 13349-13354 (1984).
5. Marchildon, G. A., Casnellie, J. E., Walsh, K. A. & Krebs, E. G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7679-7682 (1984).
6. Shultz, A. M., Henderson, L. E., Oroszlan, S., Garber, E. A. & Hanafusa, H. *Science* **227**, 427-429 (1985).
7. Chow, M. et al. *Nature* **327**, 482-486 (1987).
8. Schultz, A. M. et al. *Biochem. biophys. Res. Commun.* **146**, 1234-1239 (1987).
9. Buss, J. E., Mumby, S. M., Casey, P. J., Gilman, A. G. & Sefton, B. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7493-7497 (1987).
10. Tanabe, T. et al. *Nature* **315**, 242-245 (1985).
11. Medynski, D. C. et al. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4311-4315 (1985).
12. Yatsunami, K. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4316-4320 (1985).
13. Duronio, R. J., Rudnick, D. A., Adams, S. P., Towler, D. A. & Gordon, J. I. *J. biol. Chem.* **266**, 10498-10504 (1991).
14. Mumby, S. M., Heukeroth, R. O., Gordon, J. I. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 728-732 (1990).
15. Hurlley, J. B., Simon, M. I., Teplow, D. B., Robishaw, J. D. & Gilman, A. G. *Science* **226**, 860-862 (1984).
16. Tomer, K. B., Crow, F. W. & Gross, M. L. *J. Am. chem. Soc.* **105**, 5487-5488 (1983).
17. Navon, S. E. & Fung, B. K.-K. *J. biol. Chem.* **262**, 15746-15751 (1987).
18. Neer, E. J., Pulsifer, L. & Wolf, L. G. *J. biol. Chem.* **263**, 8996-9000 (1988).
19. Journot, L., Pantaloni, C., Bockaert, J. & Audigier, Y. *J. biol. Chem.* **266**, 9009-9015 (1991).
20. Denker, B. M., Neer, E. J. & Schmidt, C. J. *J. biol. Chem.* **267**, 6272-6277 (1992).
21. Watkins, P. A. et al. *J. biol. Chem.* **260**, 13478-13482 (1985).
22. Linder, M. E. et al. *J. biol. Chem.* **266**, 4654-4659 (1991).
23. Vuong, T. M., Chabre, M. & Stryer, L. *Nature* **311**, 659-661 (1984).
24. Heukeroth, R. O., Glaser, L. & Gordon, J. I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8795-8799 (1988).
25. Kishore, N. S. et al. *J. biol. Chem.* **266**, 8835-8855 (1991).
26. Towler, D. A. et al. *J. biol. Chem.* **263**, 1784-1790 (1988).
27. Lochrie, M. A., Hurlley, J. B. & Simon, M. I. *Science* **228**, 96-99 (1985).
28. Fukada, Y., Ohguro, H., Saito, T., Yoshizawa, T. & Akino, T. *J. biol. Chem.* **264**, 5937-5943 (1989).
29. Fukada, Y. et al. *Nature* **346**, 658-660 (1990).

ACKNOWLEDGEMENTS. We thank M. Murata, T. Matsuda, K. Sanada (Kyoto University) and J. Tamura and B. D. Musselman (JEOL USA Inc.) for their assistance in the experiments. This work was supported partly by a research grant from the Human Frontier Science Program and by Grants-in-Aid from the Japanese Ministry of Education, Science and Culture.

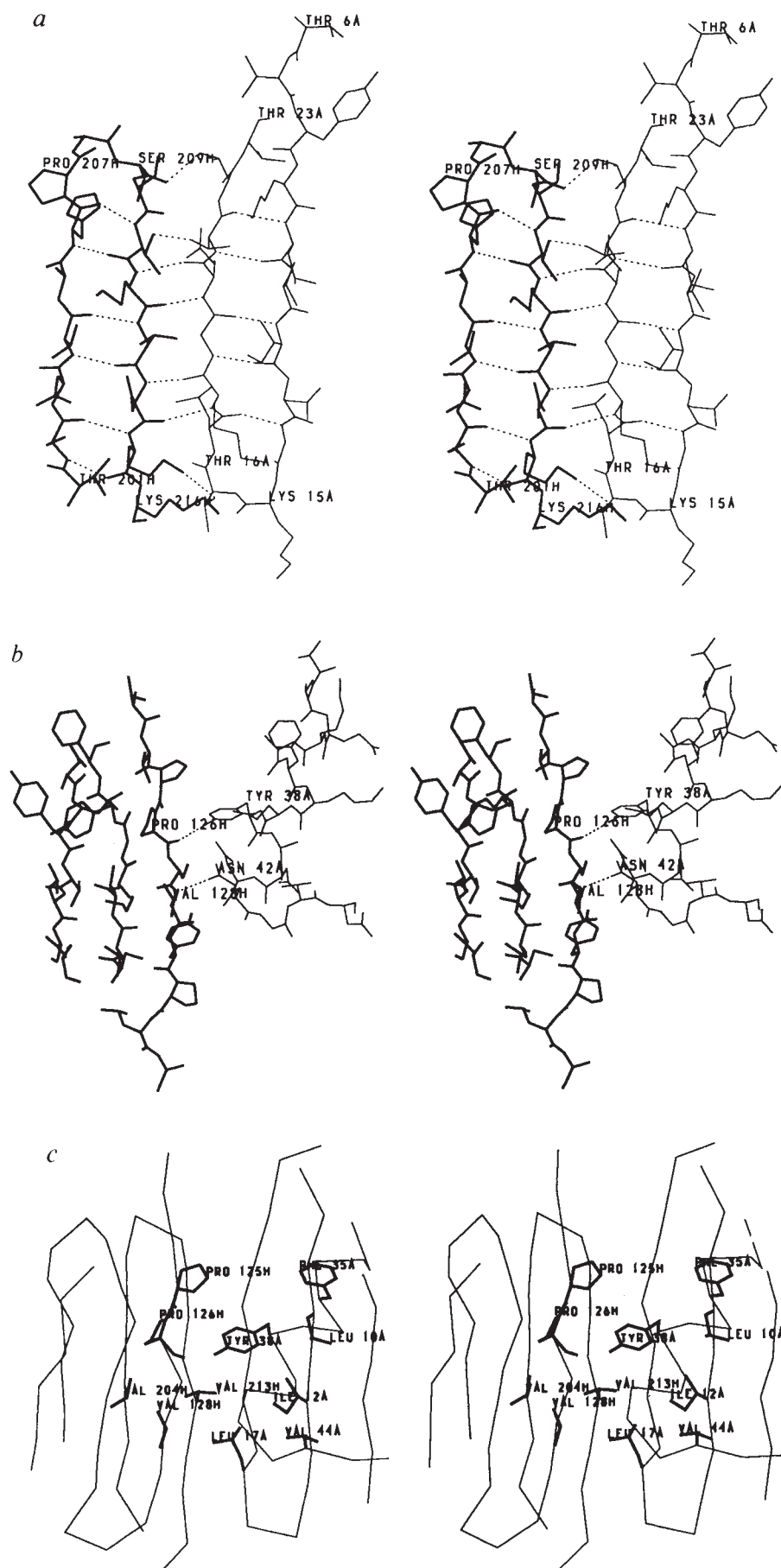


FIG. 2 Structural details of the complex. Heavy chain residue numbers are followed by H and protein G residues by A. *a*, Stereoscopic view of the hydrogen-bonds (dashed lines) formed by the last two β -strands in the CH1 domain (thick lines) and the first two β -strands in protein G (thin lines). Selected residues are indicated at the beginning and end of each strand. *b*, Stereoscopic view of the interaction between the C terminal end of the α -helix in protein G (thin lines) and the first β -strand in the CH1 domain (thick lines). Residues involved in the two hydrogen-bonds are marked. *c*, Stereoscopic view of the hydrophobic core in the complex. Side chains are shown for some of the residues that contribute to this core, superimposed on the framework of $C\alpha$ atoms.

TABLE 1 Data collection and refinement

Spacegroup $P2_12_12_1$, $a = 64.5$ Å, $b = 70.5$ Å, $c = 120.1$ Å				
Statistics of data collection				
Resolution (Å)	N_{obs}	N_{unique}	R_{merge}	Completeness
20–2.6	51,377	17,204	7.9%	98.7%
Final refinement parameters				
No. of atoms (non-hydrogen)			3,778	
No. of water molecules			279	
Resolution (Å)			10–2.6	
R -factor (final) all data			20.9%	
R.m.s. deviations from ideal geometry of the final model				
Distances (Å)				
Bonds (1–2 neighbour)			0.017	
Angles (1–3 neighbour)			0.065	
Interplanar (1–4 neighbour)			0.080	
Planar groups (Å)			0.014	
Chiral volumes (Å ³)			0.147	

Data were collected from a single crystal on beam line X-31 at the EMBL outstation at Hamburg using the image plate scanner and were processed using MOSFLM (Imperial College, London) and the CCP4 program suite (Daresbury, UK). The structure was solved by molecular replacement using CCP4 programs. The solution was found initially using the entire Fab molecule, and confirmed using two half-molecules comprising the constant and variable regions. Rigid-body refinement (as implemented in X-plor) was used to define precisely the position and orientation of each separate immunoglobulin domain. The R -factor at this stage was 48.3%, using all data between 10 and 3 Å resolution. At this stage the model was modified to the correct amino-acid sequence. Initial refinement used Hendrickson–Konnert least-squares minimization¹⁸. A few rounds of model building using FRODO¹⁹, followed by cycles of least-squares refinement reduced the R -factor to 33%. At this stage, density corresponding to protein G was clearly visible. The model for protein G used to fit the density was derived from the structure of the same domain solved at 1.1 Å resolution (J.P.D. and D.B.W., manuscript in preparation). The orientation and position of the model was refined as a rigid body in X-plor; after this procedure the R -factor had fallen to 30%. A cycle of molecular dynamics refinement reduced the R -factor to 23.5%, after which further rounds of model building and least-squares refinement were applied to insert solvent molecules.

lysozyme coordinates removed. Refinement of the structure used a combination of molecular dynamics (X-plor)^{6,7} and least-squares minimization; the final R -factor was 20.9% for all data between 10 and 2.6 Å resolution (Table 1). Density corresponding to the protein G domain appeared during the refinement adjacent to the CH1 domain in the immunoglobulin. In the final structure it is clear that protein G binds exclusively to this domain and there are no direct contacts between protein G and either the light chain or the variable region of the heavy chain (Fig. 1). The CH1 domain is relatively invariant between different Fabs derived from IgG and therefore represents the most attractive binding site for protein G.

The interaction between protein G and CH1 is accomplished by an alignment of β -strands in the two proteins, extending the β -sheet from the immunoglobulin domain into protein G

(Fig. 2). Protein G has an α/β structure in solution^{8,9}, which is preserved on binding to Fab. The complex is stabilized through a network of hydrogen-bonds extending from the last β -strand in the CH1 domain (residues 209 to 216) to the second β -strand in protein G (residues 16 to 22). This includes a pattern of hydrogen bonding at the centre of this region which is typical of antiparallel β -sheet, from residues 212 to 214 in the heavy chain and 18 to 20 in protein G. In addition, there are also hydrogen-bonds from the side chains of threonines 16, 21 and 22 in protein G to heavy chain residues. A second, minor region of contact occurs between the C-terminal end of the α -helix in protein G (residues 42 and 38) and the first β -strand of the CH1 domain (residues 128 and 126). The atoms that are involved in these interactions are predominantly from the heavy chain backbone, but use both main chain and side chain atoms from protein G. The complex is stabilized further by the association of exposed nonpolar residues from Fab and protein G, providing a continuous hydrophobic core which is shielded from solvent. All the side chains from the heavy chain involved in hydrogen bonding or hydrophobic interactions are highly conserved in all four gamma heavy chain subclasses in mouse and human.

The nature of the interactions between the proteins in the complex provides a possible explanation for the wide specificity of protein G for IgG. First, this is achieved by recognition of the CH1 domain, the portion of Fab that is most highly conserved. Second, the recognition of Fab by protein G is mediated largely by backbone atoms in the immunoglobulin, a feature which minimizes the effects of sequence variability. In cases where side chain atoms from the heavy chain are involved, these residues are highly conserved. Every aspect of the binding of protein G to Fab has been optimized to extend the range of IgG subtypes that are recognized. This has important implications for the study of protein–protein interactions¹⁰ because it demonstrates how one protein can bind to all members of a class of highly variable proteins by means of secondary structural interactions and recognition of highly conserved residues. The model for the protein G/Fab complex contrasts with the crystal structure of *Staphylococcus* protein A bound to human Fc, where contacts are mediated by side chain/side chain interactions between protein A and Fc (ref. 11). Protein G also binds to Fc with high affinity^{3,12}, and it will be particularly interesting to determine whether this interaction is mediated by association of β -strands, or through the α -helix in an analogous fashion to protein A.

We anticipate that the association between protein G and the CH1 domain is unlikely to be affected by the binding of antigen to Fab (refs 13, 14), so that protein G can be used to mediate the labelling of Fab fragments with a variety of fluorophores and detection enzymes in enzyme-linked immunosorbent assay procedures. Recent developments in the production and antigen-directed selection of recombinant Fab fragments from *Escherichia coli*^{15–17} indicate that the Fab-binding properties of protein G could find extensive application in the future. □

Received 24 June; accepted 14 September 1992.

1. Björck, L. & Åkerström, B. in *Bacterial Immunoglobulin-Binding Proteins* (Boyle, M. D. P., ed.) 113–126 (Academic, New York, 1990).
2. Björck, L. & Kronvall, G. *J. Immun.* **133**, 969–973 (1984).
3. Fahnestock, S. R., Alexander, P., Filipula, D. & Nagle, J. in *Bacterial Immunoglobulin-Binding Proteins* (ed. Boyle, M. D. P.) 133–148 (Academic, New York, 1990).
4. Myhre, E. B. in *Bacterial Immunoglobulin-Binding Proteins* (ed. Boyle, M. D. P.) 243–256 (Academic, New York, 1990).
5. Derrick, J. P., Davies, G. J., Dauter, Z., Wilson, K. S. & Wigley, D. B. *J. molec. Biol.* **227**, 1253–1254 (1992).
6. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
7. Brünger, A. T. *X-plor Manual* Version 2.1 (Yale Univ., 1990).
8. Lian, L.-Y. et al. *Biochemistry* **30**, 5335–5340 (1991).
9. Gronenborn, A. M. et al. *Science* **253**, 657–661 (1991).

10. Janin, J. & Chothia, C. *J. biol. Chem.* **265**, 16027–16030 (1990).
11. Deisenhofer, J. *Biochemistry* **20**, 2361–2370 (1981).
12. Stone, G. C. et al. *J. Immun.* **143**, 565–570 (1989).
13. Davies, D. R., Padlan, E. A. & Sheriff, S. A. *Rev. Biochem.* **59**, 439–473 (1990).
14. Colman, P. M. *Adv. Immun.* **43**, 99–132 (1988).
15. Hoogenboom, H. R. et al. *Nucleic Acids Res.* **19**, 4133–4137 (1991).
16. Sawyer, J. R. & Blattner, F. R. *Prot. Engng* **4**, 947–953 (1991).
17. Skerra, A. & Plückthun, A. *Prot. Engng* **4**, 971–979 (1991).
18. Hendrickson, W. A. & Konnert, J. H. in *Structure, Conformation and Evolution* Vol. 1 (ed. Srinivasan, R.) 43–57 (Pergamon, Oxford, 1981).
19. Jones, A. T. *J. appl. Crystallogr.* **A43**, 489–501 (1978).

ACKNOWLEDGEMENTS. We thank G. J. Davies, Z. Dauter and K. S. Wilson for data collection and processing. J.P.D. is a Postdoctoral Wellcome Research Fellow and D.B.W. is an SERC Advanced Fellow.

Neuronotes

Research tools for the neuroscientist featured this week include a patch-clamp analysis software package, a selection of perfusion chambers, imaging systems and an assortment of antibodies.

PNSROLL is World Precision Instruments' new **patch-clamp analysis software**, which is designed for interactive reviewing and analysis of single-channel binary data files under the control of a Microsoft mouse (*Reader Service No. 100*). Scrolling rate, direction and time compression of the record can be changed at any time. Markers can be set, whenever required, to indicate baseline levels and display lifetime values or to enable the 'spike' search mode and its analysis features. Lifetimes of channel openings and average current levels or current amplitudes can be quickly measured over sections of the trace and the results printed or stored on file, WPI says. In addition, the current records and histograms can be reviewed onscreen or plotted on a Hewlett-Packard plotter or HPLaserJet printer. PNSROLL runs on IBM or compatible computers, ideally equipped with EGA or VGA graphics cards and colour monitors, although it will operate with Hercules-compatible graphics cards. All programs have been written under Turbo C (Version 2.0) and accept data from standard binary files. A demo disk is available on request.

■ Research reagents

Calcium probes that are loaded into cells either as acetoxymethyl esters or as salts by microinjection or cell permeabilization techniques are often sequestered into intracellular compartments or bound to cytoplasmic proteins, giving rise to inaccurate determinations of intracellular calcium. To circumvent these problems, Molecular Probes offers several **calcium indicators** that have been covalently attached to water-soluble, high-molecular-weight dextrans — a molecular configuration that allows the indicator to remain uniformly distributed throughout the cytosol (*Reader Service No. 101*). Using the company's 10,000 Da fura-2 dextran, Miller and colleagues (*J. Cell Sci.* **101**, 7; 1992) have demonstrated the relationship between Ca^{2+} gradients and pattern of growth in pollen tubes of *Lilium*. The Calcium Green-1 dextrans should also be appropriate for monitoring changes in intracellular free Ca^{2+} that accompany the photolysis of diazo-2 or other caged Ca^{2+} probes. If Calcium Green-1 dextran is co-injected with a Ca^{2+} -insensitive fluorescent dextran of a similar molecular weight but of a contrasting colour (such as Texas Red dextran), Molecular Probes says that ratio emission determinations of intracellular calcium can be made. The Calcium Green-1 and fura-2 dextrans (10,000 Da) have been shown to undergo neuronal cell transport. These probes have been used to map

cytosolic calcium concentrations in growing rat sensory axons and to investigate mechanisms of activity-dependent calcium accumulation in embryonic chick motoneurons.

Maitotoxin, a **bioactive polyhydropolyether** isolated from the marine dinoflagellate, *Gambierdiscus toxicus*, causes an activation of L-type calcium channels and stimulation of phosphoinositide breakdown. It is available from Kamiya Biomedical (*Reader Service No. 102*). Maitotoxin has been shown to lack ionophoretic activity in mitochondria and liposomes and its highly polar nature and high molecular weight appear to preclude passage across cell membranes. The company says that maitotoxin is one of the most potent marine toxins known, with a minimum lethal dose of 170 ng per kg (IP) in mice and an EC_{50} value of less than 1 nM for phosphoinositide breakdown. Its effect on the Ca^{2+} uptake system makes it a useful tool in the study of smooth and cardiac muscle contraction and of hormone and neurotransmitter systems.

■ Perfusion studies

Lazar Research Laboratories offers a series of **biocompatible micro flow-through electrodes**, which are intended for perfusion studies on whole organs or tissue slices (*Reader Service No. 103*). These electrodes include micro flow-through modules for pH, pO_2 , pCO_2 and temperature. Lazar says that they offer researchers on a low budget a cost-effective approach to the real-time monitoring of organ perfusion experiments. Lazar also offers the four sensors as a complete monitoring system (Model PFS-2), which includes a console to house all the sensing modules, a flow-through cell tray to simplify plumbing of the flow-through

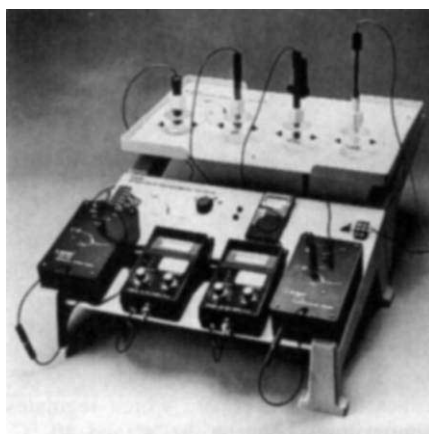
electrodes, a digital, real-time monitor and channel selector for independent monitoring of each parameter and an output connector for direct hook-up to a strip-chart recorder. The system can also be interfaced to an IBM or compatible computer. Lazar says that all sensors and components of the system are resistant to protein build-up or clogging. Sensor and flow-cell components can be chemically sterilized.

The Brandel Superfusion 2000 System, which is available from Semat, offers a **superfusion package** for pharmacological and pathophysiological research into neurotransmitter release (*Reader Service No. 104*). The system incorporates all of the essential features needed for release studies:



Superfusion 2000 system — see Semat.

a well oxygenated, temperature-controlled sample chamber to maintain the tissue preparation, a constant perfusion flow rate, a system for chemical or electrical stimulation of the tissue and a fraction collector for perfusate. The multichannel design of the Superfusion 2000 allows the simultaneous perfusion of 20 tissue samples, each with up to six different reagents in any preprogrammed sequence. Effluents can be collected for measurement of either endogenous neurotransmitter or preloaded radioisotopic labelled transmitter. Tissues are housed in individual chambers within a temperature-controlled water jacket. The volume of each chamber can be tailored to suit user requirements to between 0.2 and 0.6 ml, although larger chambers are available for synaptosomes. Electrical stimulation is provided via a variable 0–200 mA per chamber output stimulator with individual chamber current control. Tissue superfusion is provided by a specially designed, multichannel peristaltic pump. Models that are compatible with most scintillation counters and smaller, semi-automated systems (the Superfusion 600 and 1200), which offer the same perfusion technology in either a 6- or 12-channel format, are also available.



Biocompatible micro flow-through sensors.

ADVERTISEMENTS

Swant

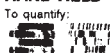
P.O. Box 2660, CH-6501 Bellinzona
Switzerland. FAX: (+41) 92 25 76 08

ANTIBODIES AGAINST EF-HAND CALCIUM BINDING PROTEINS

PARVALBUMIN CALBINDIN D-28 k CALRETININ

as well as the respective
recombinant antigens. Antibodies
excellent for immunohistochemistry,
immunoblot, radioimmunoassay etc..
Ask for complete catalogue.

Reader Service No.28

Zeinh 1D/2D/DNA/TLC
HAND-HELD SCANNING DENSITOMETER
To quantify:
 
Gels, blots, TLC, X-Ray film, Photo. 1-D pkg price \$995

8 bits, 60 um resolution, 0-3 OD, sample 11x40 cm, 18M or clones
Options: UV, 2-D, DNA, ELISA, colony counting..... \$995 each.
Universal software: for any scanner. Main 1D, 2D, DNA, \$995 each.
Representatives needed, 30% discount of USA price list.

Do-it-all, Departmental, Laser + Video
Micro Densitometry Station
\$ 16,990



BIOMED INSTRUMENTS INC. 1020 S Raymond Av., #B, Fullerton
CA 92631. Phn: (800) 927-2290, (714) 870-0290, Fax: 17148706385

Reader Service No.22

Oligos Etc. Inc.

DNA \$4.50/base
S-DNA \$8/base
Methyl Phosphonate DNA \$8/base
RNA \$8/base
S-RNA \$12/base
O-Methyl RNA \$15/base
Chimerics, Biotin
Modifiers, Fluorescent Tags etc.
No Set-Up Charges

800-888-2358/fax 800-869-0813
Intl.fax 203-438-3637

Reader Service No.30

Professional Custom Services

DNA SYNTHESIS PROTEIN SEQUENCING PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

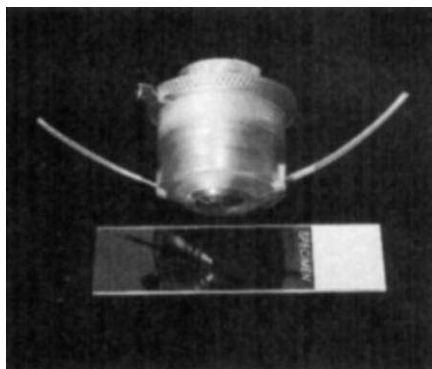
V.G. Services

1200 Wonderland Rd. S. Bldg. #7, Unit #2
London, Ontario Canada N6L 1A8

CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652 0053

Reader Service No.31

Using the CPC-2 **confocal perfusion chamber** from Adams & List researchers can perfuse living tissue with drugs or other compounds while observing them for various characteristics on confocal and standard compound microscopes (*Reader Service No. 105*). In addition, a temperature



Confocal microscope perfusion chamber.

probe can be mounted in the chamber to permit temperature regulation. The chamber consists of two components: one part attaches to a water immersion objective, the other is free to slide up and down over it. The sliding component rests on a glass slide forming the bottom of the chamber. Models are available for most water immersion lenses.

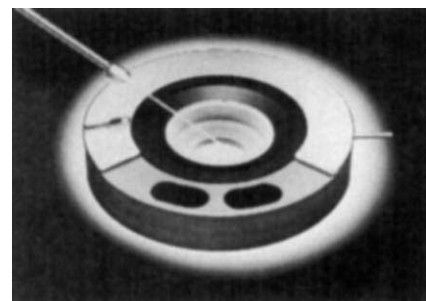
According to Medical Systems, cells can be kept viable for 10 hours or more in the company's new **brain slice chamber** (*Reader Service No. 106*). Based upon the designs of H. L. Haas and K. Zbic, the new chamber features a single base unit and two inter-



Brain slice chamber.

changeable top units. Haas' top configuration utilizes semisubmersion techniques, which are designed to provide the stability needed for intracellular recording and the rapid exchange of perfusing fluids. The Zbic top, however, is designed to maintain viable slices of tissue submerged under a constant flow of perfusing liquid. Medical Systems also offers a companion temperature controller, which can be used with either top chamber and which regulates temperatures between 30 °C and 40 °C, ± 0.2 °C.

The Model MS-200 is a complete **microincubation system** from Narishige, which consists of the Leiden microincubator and the company's digital, low-noise temperature controller (*Reader Service No. 107*). The system is recommended for studies involving electrophysiology, microinjection, time-lapse photography, phase/interference contrast microscopy and fluorescence. The Leiden microincubator provides a stable microculture environment, whether thermal, fluid or gaseous, for extended observation of cells and tissue. It

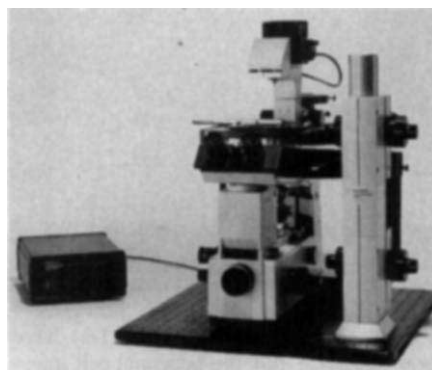


Narishige's microincubation system.

also allows convenient mechanical access for inserting measuring probes (temperature, pH) and electrodes in and out of the chamber. Two chamber options are available: either 35-mm Petri dishes or the Leiden cover slip dish — a simple, reusable cell and tissue culture dish. Narishige's microprocessor-driven temperature controller enables accurate, proportional temperature control using a thermistor that digitally controls and maintains the temperature of the chamber.

■ Imaging systems

Life science researchers can now retrofit a standard inverted fluorescence microscope and convert it into a confocal microscope with the Model VX100R **confocal microscope adaptor** from Newport/Klinger, (*Reader Service No. 108*). The company says that the system is capable of generating high-resolution images, allowing optical sectioning of cultured specimens and making electrophysiological measurements simultaneously. As the VX100R uses the arc lamp source of a conventional microscope,



Newport/Klinger's conversion kit.

it can be used with most fluorescent stains over a wide wavelength band. Preparations can be examined in real time directly through the eyepiece. Five settings provide a choice of image brightness levels for optical sectioning of weak fluorescent samples.

The new FT 1000 **spectral imaging system** from Cell Robotics, delivers multiparameter spectral imaging for the analysis of complex biological samples (*Reader Service No. 109*). Using proprietary technology, the FT 1000 converts a standard laser-scan confocal microscope into a multichannel imaging spectrometer. By performing high-speed spectral analysis, CRI says that the FT 1000 system can resolve and image up to eight spectral parameters. Simple optics and analog electronics are intended to simplify set-up, operation and maintenance. Spectral characteristics can be configured automatically using control samples or can be recalled from computer storage. The FT 1000 system does not require optical filters. This, CRI says, cuts costs, reduces light loss and simplifies operation of the instrument. In addition, as the system can be adjusted automatically for new fluorescent probes, operator intervention is kept to a minimum. Stated applications include the imaging of complex two-dimensional and three-dimensional structures, such as tissues, cells and chromosomes, where an analysis of the location of multiple fluorescent probes provides insight into the nature and possible interactions of structural elements.

Carl Zeiss announces significant upgrades to its Attofluor **digital fluorescence microscopy workstation** (*Reader Service No. 110*). Optimized for the Axiovert mi-



Carl Zeiss' real-time fluorescence image analyser for kinetic metabolic studies.

croscope and incorporating a 486/50 MHz real-time image analyser, Carl Zeiss says that the Attofluor system offers increased power, flexibility and stability in an easy-to-use, compact configuration. The image analysis workstation provides image acquisition, processing and archiving functions for the measurement of high-speed cell metabolic kinetics such as calcium, pH, cyclic AMP and other biological processes. In addition, Carl Zeiss offers Attofluor as an integrated turnkey system or as individual components. The workstation features a

highly sensitive, compact, intensified CCD camera system with multiple gain control. A vibration-free, standalone excitation filter selector facilitates image acquisition at rates of up to four ratios per second and provides simultaneous fluorescence and differential interference contrast imaging with a special CCD camera which provides the same magnification factor as the intensified CCD camera. Camera configurations are also available for studies with dual emission probes. Menu-driven software manages the wavelength selector, shutter, TV camera, image processor and video hardcopy device. Autograf software assists the user in evaluating experimental results on the Attofluor or any standalone PC.

■ Assorted antibodies

Three new **monoclonal antibodies to the endothelins**, ET-1, ET-3 and the C-terminal region of all endothelins, are now available from Advanced Biotechnologies (*Reader Service No. 111*). The monoclonals have been extensively characterized for ET-1 and ET-3 using the C-terminal monoclonal as the capture antibody in an IRMA assay sensitive to 0.06 pmol l^{-1} . Applied Biotechnologies says that crossreactivity is 10 per cent with ET-3 in the ET-1 assay and 20 per cent with ET-1 in the ET-3 assay. The C-terminal monoclonal does not show any crossreactivity with big endothelin. The monoclonals, which are IgG1 or IgG2b immunoglobulin isotypes, may also be suitable for other assay configurations, or immunocytochemistry with affinities ranging from 3.2×10^9 to $4.4 \times 10^{10} \text{ l mol}^{-1}$. Endothelin is a peptide that regulates vascular smooth muscle contraction. Increased plasma levels have been associated with a variety of disease states, including diabetes, hypertension, cardiogenic shock, acute myocardial infarction and uraemia.

Boehringer Mannheim offers three new **antibodies to the adhesion molecules** myelin-associated glycoprotein (MAG), L1 and the cell-surface glycoprotein, L-CAM/uvomorulin (*Reader Service No. 112*). Anti-MAG can be used on western blots and cryosections and reacts with MAG from several species, including mouse, frog, chicken and man. Myelin-associated glycoprotein is expressed during myelination at the contact side between axons and Schwann cells. L1 plays a role in calcium-independent cell adhesion in neurite fasciculation, migration of cerebellar granula cells and in neurite outgrowth. The antibody to L1 can be used on cryosections, fixed cells and blots and reacts with L1 from mouse and rat. Completing the line-up is the antibody to L-CAM/uvomorulin (also known as Arc-1E-cadherin or cell-CAM 120/80). L-CAM plays an important role in calcium-dependent cell-cell adhesion in epithelia irrespective of their origin from ecto-, meso- or endodermal tissue. The antibody may be used on cryosections,

ADVERTISEMENTS

MICROMANIPULATORS

Complete, easy-to-use systems for;

- Cell Isolation & Injection
- Probing ● Soldering ● Testing
- Microdissection & Microsurgery

SINGER
INSTRUMENTS
"A responsibility to Science"

Call or Fax now for further information

Tel: +44 984 40226 Fax: +44 984 41166

Singer Instrument Co Ltd., Roadwater, Watchet, Somerset, England, TA23 0QL

Reader Service No.34

CUSTOM

PEPTIDES

30 mg 20 mer
not purified

80,- DM per residue

for further information :
TIB MOLBIOL Berlin
FAX +49 30 712 7734

Reader Service No.25



Severn Biotech Ltd.

CUSTOM SERVICES

- **OLIGONUCLEOTIDE SYNTHESIS**
High quality PROBES, PRIMERS, ANTISENSE OLIGOS!
Fast-Efficient Service offering loads of modifications, BIOTIN, FLUORESCENCE, 5' & 3' labelling, Phosphothioates Affinity Support/Oligo's etc.
- **MULTIPLE PEPTIDE SYNTHESIS**
Epitope Mapping & Screening! We now supply 5mg quantities of high purity custom peptides up to 15 residues in length, using Fmoc chemistry. With D or L amino acids, these are supplied lyophilised in 4ml vials, cleaved from solid phase & ready for use.

Tel: 0562 825286 Fax: 0562 825284
Unit 23, Hookbrook Enterprise Centre
Worcester Road, Kidderminster DY10 1HY

Reader Service No.32

\$3.50/Base

Custom DNA Synthesis

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362

1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

Reader Service No.33

ADVERTISEMENTS

BIANDRM
EQUILIBRIUM DIALYSIS SYSTEM



The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIANDRM
Geräte
P.O. Box 650126
D-6800 Mannheim 85, FRG

Reader Service No.45

PACEMAKERSM

The multiple peptide synthesis and conjugation service

- High purity (up to 95%).
- Quantities for a range of uses.
- Single or multiple peptides.
- HPLC trace with each peptide.
- Terminal modification at no extra charge.
- Conjugation to carrier proteins.

from **£80 per peptide**

AFFINITY Research Products Ltd.
GPT Business Park, Technology Drive,
Nottingham, NG9 2ND, UK.
Tel: +44/0 602 436100; Fax: +44/0 602 436300

Reader Service No.23

IMAGE MASTERTM

PTI's latest Ratio Fluorescence Imaging and Microscope Photometry Systems for intracellular Ca⁺⁺, pH, Na⁺, K⁺, Mg⁺⁺ and more.

- ▶ 18 powerful IMAGEMASTER systems.
- ▶ Change/expand capabilities at any time.
- ▶ Training and technical support from PTI.

For free brochures and information contact

Tel 081 878 7928
Fax 081 876 6495

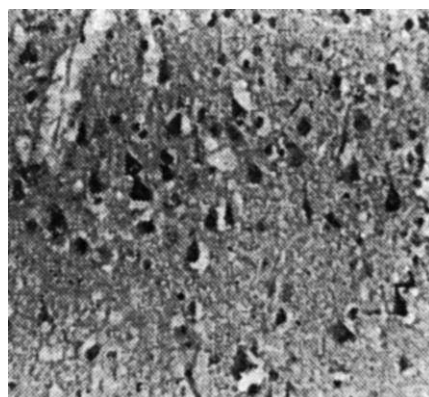
PHOTON TECHNOLOGY INT. INC
11 Well Lane, East Sheen,
London, England SW14 7AE

Reader Service No.27

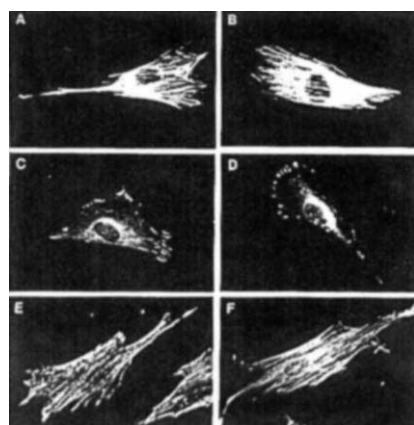
western blots and in ELISAs and reacts with L-CAM from man and rabbit. Staining of epithelial cells is confined to the lateral border, Boehringer Mannheim says.

Pierce offers over 80 new ImmunoPure **monoclonal antibodies** for use in research (Reader Service No. 113). The line-up includes monoclonals for neuroscience (fibronectin receptor, glial fibrillary acid protein), human antigens and hormones (calpastatin, carcinoembryonic antigen), human leucocytes (CD4+ helper/inducer T lymphocytes, CD8+ suppressor/cytotoxic T lymphocytes), human immunoglobulins (human IgG2 (Fab), human kappa chain) and miscellaneous antigens (bovine osteocalcin, plakoglobin).

Cymbus Bioscience is offering two **monoclonal antibodies to human neurofilaments** 160 kD and 200 kD (Reader Service No. 114). Both products can be used



Staining of tangles in pyramidal cells in the cortex reacting with Cymbus Bioscience's monoclonal antibody to human neurofilament (160 kD).



Human foreskin fibroblasts fixed with methanol and stained with a, anti-filamin (also works on paraffin-embedded tissue), b, anti-filamin (90-kD epitope), c, anti-talin, d, anti-vinculin (also works on paraffin-embedded tissue), e, anti-α actinin and f, anti-caldesmon.

in western blot studies and the detection of neurofilament antigens in frozen sections and formalin-fixed, paraffin-embedded tissues. Cymbus Bioscience says that the antibodies will recognize many neurites in human senile plaques and neurofibrillary tangles. The company also offers monoclonal antibodies to a range of attachment proteins, including filamin, 90-kD epitope of filamin, talin, vinculin, α actinin and caldesmon.

Autoantibody assessment

Bühlmann Laboratories has developed an **ELISA kit to determine autoantibodies against ganglioside M1**, an acidic glycosphingolipid of neuronal membranes in the central and peripheral nervous system (Reader Service No. 115). Antiganglioside M1 has been proven to be a useful diagnostic marker in patients with lower motor neuron disorders. As high titres of antiganglioside M1 are rare in patients with amyotrophic lateral sclerosis (ALS), this marker can be used to distinguish lower motor neuron disorder syndromes showing similar symptoms to ALS. Unlike ALS, multifocal motor neuropathy can be treated and so a determination of antiganglioside M1 levels by ELISA may help in disease assessment and treatment. Bühlmann says that the antiganglioside M1 ELISA is the first in a whole series of neuroimmunological reagents being developed by the company.

Autozyme from Cambridge Life Sciences allows **simultaneous screening of up to seven autoantibodies** of clinical significance and can be used as an aid to the diagnosis of rheumatic conditions (Reader Service No. 116). Providing results in less than two hours, Autozyme enables researchers to screen simultaneously for Ro, La, Sm, Sm/RNP, Jo-1, Scl-70 and nDNA autoantibodies in human serum or plasma. The addition of nDNA to the range of extractable nuclear antibodies covered by Autozyme is intended to provide a simple and cost-effective means of screening for this analyte. Autozyme uses antigen-coated microwell technology and a noncompetitive enzyme immunoassay method that is identical for all seven antigens. All the assays use affinity-purified antigens. Using the appropriate combination of assays and a standardized assay protocol, Autozyme allows researchers to identify or confirm a range of conditions, including Sjogren's syndrome, systemic lupus erythematosus, mixed connective tissue disease, polymyositis and systemic sclerosis. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.